JOURNAL OF EDUCATIONAL RESEARCH

• • •

EDITORIAL BOARD

A. S. BARR
Professor of Education, University of Wisconsin, Madison, Wisconsin

WILLIAM A. BROWNELL, Dean, School of Education, University of California, Berkeley, California.

ROBERT A. DAVIS, Professor of Educational Research, George Peabody College for Teachers, Nashville, Tennessee.

GUY D. McGRATH, Dean, College of Education, Arizona State University, Tempe, Arizona.

HAROLD E. MITZEL, Director, Office of Research and Evaluation, Division of Teacher Education, College of the City of New York, New York City.

WILLARD C. OLSON, Dean of College of Education, University of Michigan, Ann Arbor, Michigan.

JOHN SCHMID, Professor, College of Education, University of Arkansas, Fayetteville, Arkansas.

WILSON B. THIEDE, Professor of Education, University of Wisconsin, Madison, Wisconsin.

ARTHUR E. TRAXLER, Executive Director, Educational Records Bureau, New York, New York.

VOLUME LIV
September 1960 - May 1961

BUREAU OF EDNL. & PSYL.
LIBRARY

Copyright 1960-1961 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

Bur. Edn. Research
1·9·70
J 300

INDEX

Listed Under Articles and Authors

Aaron, Ira E., Comparisons of Good and Poor Readers in Fourth and Eighth Grades, September, 1960, 34-37.
A Beginning Vocabulary, Daniel W. Fullmer et al., March, 1961, 270-272.
Acker, Ralph S., Reading Improvement in Military, Government and Business Agencies, February, 1961, 238-240.
A Comparison of the WISC Patterns of Male Retarded and Non-Retarded Readers, Donald Neville, January, 1961, 195-197.
Actuarial Prediction in Guidance, Henry D. Rinsland, January, 1961, 168-172.
Adequacy and Usefulness of an Objective Language Scale When Administered to Elementary School Children, William G. Williams, September, 1960, 30-33.
A Follow-up Study of Teacher Candidates, J. C. Gowan et al., May 61, 353-355.
Aliferis Music Achievement Test as a Predictor of Success in Music Theory at St. Olaf College, Adolph White, April, 1961, 315-317.
A Longitudinal Study of the Relationship of High School Foreign Language and Mathematics Study to Freshman Grades, Joseph V. West et al., November, 1960, 105-110.
Alternative Course in Secondary Mathematics, Daniel S. Parkinson, December, 1960, 128-132.
Analysis of Attitude Inventory Data on Minnesota Vocational Agriculture Teachers, Gordon I. Swanson et al., May, 1961, 349-352.
An Attention Scale for Evaluating E.T.V. Programs, Cecil C. Bridges, Jr., December, 1960, 149-152.
Anderson, A. E., Personality Traits in Reading Ability of Western Australian University Freshmen, February, 1961, 234-237.
Anderson, C. C., One Explanation of the Extent of Function Fluctuation Observed in Total and Subtest Cognitive Performance, December, 1960, 123-127.
An Experimental Study of the Self-Concept of Student Teachers, Agnes B. Hatfield, May, 1961, 356-358.
An Experiment With the Controlled Reader, Forbes Bottomly, March, 1961, 265-269.
A Note on the Validity of Borow's College Inventory of Academic Adjustment, W. James Popham et al., November, 1960, 115-117.
Applbaum, Morris J., Special Budgetary Provisions Made for Educating the Gifted, December, 1960, 145-148.
A Re-evaluation of the Health Railwalking Test, Cornelius P. Goetzinger, January, 1961, 187-191.
Art Activities for Mentally Handicapped Children, Richard G. Wiggin, March, 1961, 251-258.
Assertion of Authority: Differences Between Teachers, Student-Teachers, and Mothers of Young Children, Albert S. Dryer et al., October, 1960, 63-66.
A Study of Academic Achievement and Performance on a Test of Library Understandings, William D. Joyce, January, 1961, 198-199.
A Study of Achievement in Two Types of Public Secondary Schools, John W. French, September, 1960, 25-29.
A Study of the Socio-Economic Backgrounds of Parents of Miami University (Ohio) Students, Joseph C. Seibert et al., March, 1961, 259-264.
A Survey of Educational Disability in Emotionally Disturbed Children, Arthur S. Tamkin, October, 1960, 67-69.

Baker, Janet Bassett, College Students and Their Assignments, October, 1960, 49-53.
Bloomer, Richard H., Concepts of Meaning and the Reading and Spelling Difficulty of Words, January, 1961, 178-182.
Borg, Walter R. et al., Four Criteria of Principal Effectiveness, May, 1961, 332-337.
Borgatta, Edgar F. et al., The Quick Word Test, September, 1960, 15-19
Bottomly, Forbes, An Experiment With the Controlled Reader, March, 1961, 265-269.
Bridges, Cecil C., Jr., An Attention Scale for Evaluating E.T.V. Programs, December, 1960, 149-152.
Brown, George I., The Relationships Between Categories of Behavior of Third Grade Teachers, May, 1961, 338-344.

Cattell, Raymond B. et al., Rationale of Norms on an Adult Personality Test, the 16 P. F. — For American Women, April, 1961, 285-290.
Characteristics of Slow Learners and Pupils of Normal Intelligence Referred to a Child Guidance Clinic, Walter I. Murray et al., October, 1960, 43-48.
Chase, Clinton I., The Position of Certain Variables in the Prediction of Problem-Solving in Arithmetic, September, 1960, 9-14.
Children's Likes and Dislikes About What They Read, J. H. Coleman et al., February, 1961, 221-228.
Cole, David L., The Prediction of Teaching Performance, May, 1961, 345-348.
Coleman, J. J., et al., Children's Likes and Dislikes About What They Read, February, 1961, 221-228.
College Students and Their Assignments, Janet Bassett Baker, October, 1960, 49-53.
Comparison of Education Students, Successful Science Teachers and Educational Administrators on the Edwards PPS, Reed M. Merrill, September, 1960, 38-40.
Comparisons of Good and Poor Readers in Fourth and Eighth Grades, Ira E. Aaron, September,

1960, 34-37.
Concepts of Meaning and the Reading and Spelling Difficulty of Words, Richard H. Bloomer, January, 1961, 178-182.
Confidential Information and the Guidance Program, Arthur A. Rezny et al., March, 1961, 243-250.
Counselor and Student Estimates of Grades as Predictors of Academic Achievement, Malcolm H. Robertson, October, 1960, 73-75.
Criteria for Industry Sponsored Educational Material and Selected Sources for Chemistry, Keith I. Roper et al., November, 1960, 99-104.

Determining Individual Needs of Elementary Teachers as a Basis for an Orientation Program, Anne Phillips McCreary, September, 1960, 20-24.
Development of the Ideal Self in Western Australian Youth, D. K. Wheeler, January, 1961, 163-167.
Devereux Readability Index, Edgar A. Smith, April, 1961, 298-303.
Dreyer, Albert S. et al., The Assertion of Authority: Differences Between Teachers, Student-Teachers, and Mothers of Young Children, October, 1960, 63-66.

Effect of Training in Listening for Specific Purposes, Sue E. Trivette, March, 1961, 276-277.
Effects of the Purported Validity of a Test Upon Level of Motivation and Performance, Lawrence S. Wrightsman, Jr., December, 1960, 153-156.
Evidence from School Records on the Effectiveness of Ability Grouping, John W. French, November, 1960, 83-91.
Experiment in Cross Grouping, Jeff West et al., October, 1960, 70-72.

Factor Analyses of High School Science Achievement Measures, Leroy Wolins et al., January, 1961, 173-177.
Faculty and Student Perceptions of Cheating, Jack R. Frymier, November, 1960, 118-120.
Fleming, Thomas F., Further Appraisal of the Manner in Which the American Press Handles Material Relevant to Education, November, 1960, 92-98.
Four Criteria of Principal Effectiveness, Walter R. Borg et al., May, 1961, 332-337.
French, John W., A Study of Achievement in Two Types of Public Secondary Schools, September, 1960, 25-29
French, John W., Evidence from School Records on the Effectiveness of Ability Grouping, November, 1960, 83-91.
Frymier, Jack R., Faculty and Student Perceptions of Cheating, November, 1960, 118-120.
Fullmer, Daniel W. et al., A Beginning Reading Vocabulary, March, 1961, 270-272.
Further Appraisal of the Manner in Which the American Press Handles Material Relevant to Education, Thomas F. Fleming, November, 1960, 92-98.

Gekoski, Norman et al., Student Mortality and Related Factors, January, 1961, 192-194.
Goetzinger, Cornelius P., A Re-evaluation of the Health Railwalking Test, January, 1961, 187-191.
Gowan, J. C. et al., A Follow-up Study of Teaching Candidates, May, 1961, 353-355.
Gowan, J. C., Intercorrelations of the American Council Psychological Examination with Other Types of Tests, December, 1960, 157-159.
Guggenheim, Fred, Nonteaching Activities of Teachers in a Staff Utilization Study, May, 1961, 325-331.

Halliwell, Joseph W., Relationship of Certain Factors to Marking Practices in Individualized Reporting Programs, October, 1960, 76-78.
Hatfield, Agnes B., An Experimental Study of the Self-Concept of Student Teachers, May, 1961, 356-358.

Incidental Learning in Foreign Language Learning, Paul Pimsleur, November, 1960, 111-114.
Influences on Students' Perception and Persistence in the Undergraduate College, J. M. Slater, September, 1960, 3-8.
Intercorrelations of the American Council Psychological Examination with Other Types of Tests, J. C. Gowan, December, 1960, 157-159.

Joyce, William D., A Study of Academic Achievement and Performances on a Test of Library Understandings, January, 1961, 198-199.

Markert, Marlow A., Social Status and Enrollment in Traditional Mathematics, October, 1960, 79.
McAulay, J. D., What Understandings Do Second Grade Children Have of Time Relationships? April, 1961, 312, 314.
McCreary, Anne Phillips, Determining Individual Needs of Elementary Teachers as a Basis for an Orientation Program, September, 1960, 20-24.
McKenna, David L., Toward a Theory of Interpersonal Relationships in the Administration of Higher Education, December, 1960, 133-136.
Men Teachers in Public Elementary Schools, Robert S. Thurman, October, 1960, 54-58.
Merrill, Reed M., Comparison of Education Students, Successful Science Teachers and Educational Administrators on the Edwards PPS, September, 1960, 38-40.
Missing Data in Non-Linear Trend Analysis of Repeated Measurements on the Same Individuals, Richard B. Seymour, December, 1960, 141-144.
Murray, Walter I., et al., Characteristics of Slow Learners and Pupils of Normal Intelligence Referred to a Child Guidance Clinic, October,

1960, 43-48.

Natchez, Gladys, Oral Reading Used as an Indicator of Reactions to Frustration, April, 1961, 308-311.
Neville, Donald, A Comparison of the WISC Patterns of Male Retarded and Non-Retarded Readers, January, 1961, 195-197.
Nonteaching Activities of Teachers in a Staff Utilization Study, Fred Guggenheim, May, 1961, 325-331.

One Explanation of the Extent of Function Fluctuation Observed in Total and Subtest Cognitive Performance, C. C. Anderson, December, 1960, 123-127.
Oral Reading Used as an Indicator of Reactions to Frustration, Gladys Natchez, April, 1961, 308-311.

Parkinson, Daniel S., The Alternative Course in Secondary Mathematics, December, 1960, 128-132.
Parsley, K. M., Jr., et al., Relationships Between the Lee-Clark Reading Readiness Test and the 1947 Revision of the Stanford-Binet Intelligence Test, Form L, April, 1961, 304-307.
Peer Status and Reading Achievement, O. V. Porterfield et al., April, 1961, 291-297.
Personality Traits in Reading Ability of Western Australian University Freshmen, A. W. Anderson, February, 1961, 234-237.
Pimsleur, Paul, Incidental Learning in Foreign Language Learning, November, 1960, 111-114.
Popham, W. James et al., A Note on the Validity of Borow's College Inventory of Academic Adjustment, November, 1960, 115-117.
Porterfield, O. V. et al., Peer Status and Reading Achievement, April, 1961, 291-297.
Position of Certain Variables in the Prediction of Problem-Solving in Arithmetic, Clinton I. Chase, September, 1960, 9-14.
Powell, Marvin, et al., The Relationships Between First Grade Reading Readiness and Second Grade Reading Achievement, February, 1961, 229-233.
Prediction of Teaching Performance, David L. Cole, May, 1961, 345-348.

Quick Word Test, Edgar F. Borgatta et al., September, 1961, 15-19.

Rationale of Norms on an Adult Personality Test, the 16 P. F. --For American Women, Raymond B. Cattell et al., April, 1961, 285-290.
Reading Improvement in Military, Government and Business Agencies, Ralph S. Acker, February, 1961, 238-240.
Relationship Between Science Proficiency and Training Among Liberal Arts Graduates, Jacob Tuckman et al., March, 1961, 273-275.
Relationship of Certain Factors to Marking Practices in Individualized Reporting Programs, Joseph W. Halliwell, October, 1960, 76-78.
Relationships Between Categories of Behavior of Third Grade Teachers, George I. Brown, May, 1961, 338-344.
Relationships Between First Grade Reading Readiness and Second Grade Reading Achievement, Marvin Powell et al., February, 1961, 229-233.
Relationships Between the Lee-Clark Reading Readiness Test and the 1947 Revision of the Stanford-Binet Intelligence Test, Form L, K. M. Parsley, Jr. et al., April, 1961, 304-307.
Rezny, Arthur A. et al., Confidential Information and the Guidance Program, March, 1961, 243-250.
Rinsland, Henry D., Actuarial Prediction in Guidance, January, 1961, 168-172.
Robertson, Malcolm H., Counselor and Student Estimates of Grades as Predictors of Academic Achievement, October, 1960, 73-75.
Robinson, Helen M., Summary of Investigations Relating to Reading, July 1, 1959 to June 30, 1960, February, 1961, 203-220.
Roper, Keith I. et al., Criteria for Industry Sponsored Educational Materials and Selected Sources for Chemistry, November, 1960, 99-104.

Seibert, Joseph C. et al., A Study of the Socio-Economic Backgrounds of Parents of Miami University (Ohio) Students, March, 1961, 259-264.
Selected Out-of-School Factors That Affect Negro Elementary School Children, Elsie Hill Wallace, December, 1960, 137-140.
Seymour, Richard B., Missing Data in Non-Linear Trend Analysis of Repeated Measurements on the Same Individuals, December, 1960, 141-144.
Slater, J. M., Influence on Students' Perception and Persistence in the Undergraduate College, September, 1960, 3-8.
Smith, Edgar A., Devereux Readability Index, April, 1961, 298-303.
Social Status and Enrollment in Traditional Mathematics, Marlow A. Markert, October, 1960, 79.
Special Budgetary Provision Made for Educating the Gifted, Morris J. Applbaum, December, 1960, 145-148.
Student Mortality and Related Factors, Norman Gekoski et al., January, 1961, 192-194.
Summary of Investigations Relating to Reading, July 1, 1959 to June 30, 1960, Helen M. Robinson, February, 1961, 203-220.
Swanson, Gordon I. et al., Analysis of Attitude Inventory Data on Minnesota Vocational Agriculture Teachers, May, 1961, 349-352.

Tamkin, Arthur S., A Survey of Educational Disability in Emotionally Disturbed Children, October, 1960, 67-69.

Thurman, Robert S., Men Teachers in Public Elementary Schools, October, 1960, 54-58.

Toward a Theory of Interpersonal Relationships in the Administration of Higher Education, David L. McKenna, December, 1960, 133-136.

Traxler, Arthur E., Use of Results of Large-Scale Testing Programs in Instruction and Guidance, October, 1960, 59-62.

Trivette, Sue E., The Effect of Training in Listening for Specific Purposes, March, 1961, 276-277.

Tuckman, Jacob et al., The Relationship Between Science Proficiency and Training Among Liberal Arts Graduates, March, 1961, 273-275.

Use of Results of Large-Scale Testing Programs in Instruction and Guidance, Arthur E. Traxler, October, 1960, 59-62.

Values and Achievement in a Freshman Psychology Course, William D. Winter, January, 1961, 183-186.

Wallace, Elsie Hill, Selected Out-of-School Factors That Affect Negro Elementary School Children, December, 1960, 137-140.

West, Jeff et al., Experiment in Cross Grouping, October, 1960, 70-72.

West, Joseph V. et al., A Longitudinal Study of the Relationship of High School Foreign Language and Mathematics Study to Freshman Grades, November, 1960, 105-110.

What Understandings Do Second Grade Children Have of Time Relationships? J. D. McAulay, April, 1961, 312-314.

Wheeler, D. K., Development of the Ideal Self in Western Australian Youth, January, 1961, 163-167.

White, Adolph, The Aliferis Music Achievement Test as a Predictor of Success in Music Theory at St. Olaf College, April, 1961, 315-317.

Wiggin, Richard G., Art Activities for Mentally Handicapped Children, March, 1961, 251-258.

Williams, William G., The Adequacy and Usefulness of an Objective Language Scale When Administered to Elementary School Children, September, 1960, 30-33.

Winter, William D., Values and Achievement in a Freshman Psychology Course, January, 1961, 183-186.

Wolins, Leroy et al., Factor Analyses of High School Science Achievement Measures, Janary, 1961, 173-177.

Wrightsman, Lawrence S., Jr., The Effects of the Purported Validity of a Test Upon Level of Motivation and Performance, December, 1960, 153-156.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 SEPTEMBER, 1960 NUMBER 1

CONTENTS

Influences on Students' Perception and Persistence in the Undergraduate College—J. M. Slater ... 3

The Position of Certain Variables in the Prediction of Problem-Solving in Arithmetic—Clinton I. Chase ... 9

✓ The Quick Word Test—Edgar F. Borgatta and Raymond J. Corsini ... 15

Determining Individual Needs of Elementary Teachers as a Basis for an Orientation Program—Anne Phillips McCreary ... 20

A Study of Achievement in Two Types of Public Secondary Schools—John W. French ... 25

✓ The Adequacy and Usefulness of an Objective Language Scale When Administered to Elementary School Children—William G. Williams ... 30

Comparisons of Good and Poor Readers in Fourth and Eighth Grades—Ira E. Aaron ... 34

Comparison of Education Students, Successful Science Teachers and Educational Administrators on the Edwards PPS—Reed M. Merrill ... 38

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

INFLUENCES ON STUDENTS' PERCEPTION AND PERSISTENCE IN THE UNDERGRADUATE COLLEGE

J. M. SLATER
University of Illinois

"I'M JUST NOT getting anywhere" is a classic statement by the student who seeks to transfer from one college to another or who seeks to terminate his college attendance. And perhaps his analysis of his desire for a change is more of an explanation than we have thought, for "getting somewhere," making progress toward personal goals, is the basis for much of that which keeps people at daily tasks for extended periods of time.

Drop-outs have become significant as a statistic related to institutional success or failure.. When John H. McNeely completed a study of College Student Mortality in 1937 (1), he chose to introduce his final report as follows:

"One of the primary essentials to an intelligent appraisement of the success of higher education is an analysis of college student mortality. By student mortality is meant the failure of students to remain in college until graduation.

"Moreover, the first step in determining the advisability of reforms in higher education is the discovery, accumulation, and interpretation of factual data on the subject. This applies to individual universities and colleges as well as higher education in general. Reorganization of educational programs, improvement in traditional methods of instruction, inauguration of new admission or graduation requirements, readjustment of collegiate environments to students, and changes of a similar nature should be undertaken only upon a basis of detailed knowledge dealing among other things with student mortality in all its phases."

Interest in the problem continues. The Association of Collegiate Registrars and Admissions Officers has recently collaborated with the U. S. Office of Education in a study of the phenomenon (4). Dael Wolfle's report on America's Resources of Specialized Talent (6) includes much that is related to attrition among college students.

Newly established procedures to implement college choice by reducing financial obstacles (e. g., the National Merit Scholarship Program and numerous state scholarships programs) are not likely to reduce interest in this question. Such programs serve to increase the alternatives from which a student may choose. As the alternatives for choice become greater, the opportunities for inappropriate choices are also increased. Furthermore, attrition takes on even greater significance when those who depart have been depended upon to assume important and unusual roles in the period ahead. All of us recognize the possibilities which abound in the presence of talented youth.

A Context for Considering Persistence and Attrition

In March, 1957, the author (2) noted a variety of approaches which have been used in the attempt to increase our understanding of student mortality. After suggesting that we should attempt to deal with the anticipated influence of numerous independent variables which are operating to create this problem, he proceeded to describe a conceptual framework for the consideration of persistence and attrition among college men. His description originated in the Lewinian tradition which proposes that a person persists in an activity so long as the activity serves the psychological needs of the individual. In this context, persistence ends, i. e., attrition takes place, when the activity takes on the psychological characteristics of repetition, that of marking time as opposed to making progress.

The report which follows describes an attempt to test some assumptions growing out of the conceptualization noted above. The approach is somewhat novel in that it proposes to gain information about attrition, i. e., drop-outs, by studying those who persist in the undergraduate college to graduation. The setting for the study was a mid-western university, namely the University of Illinois. This seemed a logical place to test predictions concerning persistence and attrition for

a. the population of the University is large enough to provide adequate samples in numerous categories;

b. registration customarily embraces two kinds of curricula, i. e., those which emphasize vocational objectives and those which do not include or infer vocational objectives;

c. a minimum of control is achieved over environmental factors such as basic expenses, opportunity for curricular election, etc., when all cases come from the same campus.

Design of the Study

The primary assumption underlying the study was that college curricula differ not only as to purpose, content, and method, but also as to the way in which these elements are perceived by different groups of students. Furthermore, since these differences exist, we should not expect that persistence rates will be similar for all groups in a given curriculum. To test the assumption, a population of male students was grouped on the basis of fathers' occupations; the presumption being that student attitudes, values, aspirations, etc., have developed from experiences which are in no small part determined by the nature of the father's employment. It was also presumed that if the conceptual framework be adequate, and if the general climate of the University and its colleges be understood, it should be possible to predict which groups will persist the longer in specified curricula within the University.[1]*

The population consisted of 1323 males who registered in one of the following colleges on the Urbana campus of the University of Illinois in September, 1950:

College of Agriculture
College of Commerce and Business Administration
College of Engineering
College of Liberal Arts and Sciences.

In the summer of 1955, official University records [2] were examined to determine whether each of these students had or had not graduated from (or was or was not presently registered in) that college of the University in which he enrolled in 1950.

Fathers' occupations were classified on the basis of categories provided in the Dictionary of Occupational Titles (3). Colleges were differentiated on the basis of their objectives as stated in the 1953-54 catalog of the University (5).

Given this information, the basic process was one of comparing differences in rates of persistence when the students were classified according to

a. the occupation of the father

b. the college in which each was enrolled in September, 1950;

and when the rate of persistence for any given group was expressed by the ratio

$$\frac{\text{Number of persistent students in category `x', July, 1955}}{\text{Total number of students in category `x', September, 1950.}}$$

Four secondary assumptions were considered. Three of these involved a prediction as to the presence or absence of intergroup differences in attrition. The fourth was to the effect that anticipated differences in persistence would not be accompanied by differences in academic aptitude as measured by the ACE test of academic aptitude.[3]

Differences in Persistence (in the Absence of Specification as to College and Curriculum)

As noted above, the basic hypothesis was to the effect that to predict persistence one should not generalize across all students in any one curriculum, but that he should be specific concerning given groups of students in a certain curriculum. The first assumption to be tested was that rates of persistence for given groups would not differ when college and curriculum were unspecified. Table I shows the rate of persistence for each of the six groups of male students--when curriculum was not considered. Table II summarizes the t ratios for differences in these rates.[4]

The prediction of no difference was not substantiated. The rate of persistence for sons from farm homes differed significantly [5] from the rate for each of the other groups[6].

Differences in Persistence in Different Kinds of Curricula (Without Regard to Home Background of the Students)

The basic assumption implied also that it is inappropriate to generalize across all curricula when considering attrition and persistence. The second test was to determine whether persistence was greater in certain curricula than in others--when home background of the students was not considered. It was predicted that, for males at the University of Illinois, those enrolled in curricula which have specific vocational objectives that can normally be achieved in four years will persist at a higher rate than will those who are enrolled in curricula that are not oriented toward some relatively specific vocational goal[7]. Table III indicates rates of persistence, difference and t ratio for students enrolled in these contrasting curricula.

The predicted difference was observed. A chi square test of independence also substantiated the sub-hypothesis that the students in the group with the higher rate of persistence would not rank higher on the ACE than would those in the group with the lower rate of persistence[8].

---

* Footnotes will be found at the end of the article.

TABLE I

FATHERS' OCCUPATIONAL CLASSIFICATION AND SONS' PERSISTENCE AT THE UNIVERSITY OF ILLINOIS*

Fathers' Occupational Classification	Sons' Rates of Persistence
Professional	.38 (215)**
Managerial	.36 (302)
Clerical and Sales	.38 (265)
Agricultural	.51 (192)
Skilled	.35 (238)

* These men were enrolled in the Colleges of Agriculture, Commerce, Engineering, and Liberal Arts and Sciences.

** Numbers in parentheses indicate total cases in the category.

TABLE II

SUMMARY OF t RATIOS FOR DIFFERENCES IN RATES OF PERSISTENCE INDICATED IN TABLE I

	Professional	Managerial	Clerical and Sales	Agricultural	Skilled
Professional		.46	.00	2.62*	.65
Managerial			.49	3.40*	.25
Clerical and Sales				2.76*	.69
Agricultural					3.42*
Skilled					

* Significant at the 5 percent level.

TABLE III

RATES OF PERSISTENCE, DIFFERENCE, AND t FOR MEN ENROLLED IN CURRICULA WITH SPECIFIC VOCATIONAL OBJECTIVES AND MEN ENROLLED IN CURRICULA WITHOUT SPECIFIC VOCATIONAL OBJECTIVES

Persistence in Vocational Curricula	Persistence in Non-Vocational Curricula	Difference in Rates of Persistence	t
.41 (901)**	.18 (188)	.23	7.07*

* Significant at the 5% point.

** Numbers in parentheses indicate total cases in the category

TABLE IV

RATES OF PERSISTENCE, DIFFERENCE, AND t FOR "SPECIFIED" AND "OTHER" SONS IN THE COLLEGES OF AGRICULTURE, COMMERCE, ENGINEERING AND LIBERAL ARTS AND SCIENCES

College	Persistence-"Specified" Sons	Persistence-"Other" Sons	Difference	t
Agriculture	.58 (118)**	.31 (108)	.27	4.16*
Engineering	.54 (39)	.39 (324)	.15	1.89*
Commerce	.45 (96)	.35 (216)	.10	1.68*
Liberal Arts and Sciences	.44 (50)	.31 (372)	.13	1.82*

* Significant at the 5% point.

** Numbers in parentheses indicate total cases in the category.

Differences in Persistence (When Both Students and Curricula Were Specified)

The final test of the basic assumption called for a prediction as to which men in a given college would persist at a higher rate than would other men in the same college. Proceeding with the logic which related fathers' occupations and students' aspirations, values, goals, etc., it was predicted that persistence would be greater for students enrolled in colleges which prepared them to enter occupations similar to those in which their fathers were employed than for others in the same college whose fathers were otherwise employed. For example, the prediction was that students from farm homes would persist at a higher rate in the College of Agriculture than would non-farm students; that sons of engineers, chemists, and draftsmen would persist at a higher rate in the College of Engineering than would "other" sons in this college; and so also for specified groups of students in the College of Commerce and the College of Liberal Arts and Sciences.

Colleges and Fathers' Occupations were matched as follows:

College	Fathers' Occupational Groups Specified
Agriculture	Farming
Commerce	Accountant, Clerical, Sales
Engineering	Chemist, Engineer, Draftsman
Liberal Arts and Sciences	Lawyers, Teachers (secondary school and college), Authors, Editors and Reporters; Clergymen, Librarians, Curators, Physicians and Surgeons, Statisticians, Natural Scientists, Social Scientists.

As is shown by Table IV, the hypothesis as to higher rates of persistence for specified students in each college was supported. Chi square tests again supported the assumption that intergroup differences in persistence could not be explained by differences in ACE scores.

Summary

This study proceeded upon the hypothesis that the value systems and experiences which characterize family units predispose the children favorably or unfavorably toward the academic aspects of their collegiate life. When fathers' occupations were used to establish categories on the cultural criterion, and when stated objectives were used to differentiate undergraduate curricula, data on persistence rates for male students at the University of Illinois seemed to support this basic assumption.

When persistence was considered without regard for curricular differences, persistence rates were of the magnitude of .35 and .38 with one exception. The rate for all men in college with well-defined vocational objectives was .41. When students were enrolled in colleges which prepared them to enter occupations similar to those in which their fathers were employed, persistence rates ranged from .44 to .58. Persistence rates were of the magnitude of .39 and below for students in "strange" academic settings.

On the strength of the information in Tables I and II, one might conclude that men from farm homes were more persistent than all other groups of men included in this study. Actually, however, farmers' sons had concentrated their registration in the College of Agriculture where their persistence was high, thus making the rate for all farm sons appear high in contrast to other groups whose registrations were spread more evenly across all colleges. (Further analysis of the data revealed that the persistence of farm sons in "strange" curricula was as follows: in Liberal Arts and Sciences, .30; in Commerce, .29; and in Engineering, .51.)

This study was not designed to discover which elements cause a given undergraduate curriculum to be more meaningful to some students than to others but it does seem apparent that such elements exist. The results suggest one thing these elements are not. They are not of the kind one would assume if he were to select students for persistence on the basis of ACE scores. That is to say he influence of the determinants--father's occupation and vocational objectives in the curricula-- seemed not to be dependent upon ACE scores which fell in the range reported in this study. Subsequent research may identify crucial elements in the subtleties of vocabulary, procedure, objectives, etc., which are peculiar to each curriculum. It seems quite probable that daily assignments which seem worthwhile to one student are without meaning to another, and that faculty expectations which some students take for granted will be wholly unacceptable to others.

Discussion

This study supports the assumption that the incorporation of institutional, curricular and professional objectives into self is a central task which occupies those who persist in higher education. Personality theory suggests that this process of incorporation includes only that which is perceived as having some relationship to self. As has been demonstrated herein, home and family experiences influence the process of incorporation, favorably for some, perhaps not so favorably for others. Numerous other pre-conditions such as school experiences, health, relationships with peers and other adults, etc., no doubt also have a bearing upon the facility which students must develop if they are to achieve this identification of self with college programs and college people.

Singularly important to our consideration here, however, is the fact that this is a process--something which is susceptible to change as new experiences and new relationships are introduced. Herein lies the opportunity for those who set the climate for academic endeavor. Rather than focus upon the adequacy or inadequacy of meals, housing, transportation, etc., it would appear that researchers will find it more productive to look to total curriculum to see how it may be so structured, or freed from rigidity of structure, to facilitate this process of incorporation for the greatest number of students. Furthermore, it will be well to consider carefully the characteristics of those who present ideas and values to the students who choose to study in each field.

The import of these suggestions is that institutions will find it advantageous to remain continuously alert to the relationships which exist to assist students in their efforts to encompass new personal and intellectual horizons. Students who are confident in their ability to deal with new situations, who experience a minimum of threat and fear in new settings, and who find it relatively easy to identify with new persons, goals and values may need little more than an opportunity to pursue their academic careers in any field they choose. By contrast, however, many students will be characterized by feelings of inadequacy both with regard to the development of new relationships with ideas and with people. For these the situation must contain numerous therapeutic qualities if the process of incorporation is to move forward in a manner which facilitates persistence.

FOOTNOTES

1. For this study, a persistent student was one who in July, 1955, was found a) to have graduated from, or b) to be currently registered in the college which he entered at the University of Illinois in September, 1950. A non-persistent student was one who in July, 1955, was found not a) to have graduated from, or b) to be currently registered in the college which he entered at the University of Illinois in September, 1950.
2. Original academic records were made available by the Director of the Office of Admissions and Records, University of Illinois.
3. Scores on the American Council on Education Psychological Examination, College edition, 1946, were made available by the Director of the Counseling Bureau, University of Illinois.
4. It is not sufficient merely to point to observed differences between two proportions, for such differences may be expected to occur by chance or random sampling, even if the two groups had not been subjected to different conditions. The test of the reliability of differences in sampling distributions is customarily described by the term "t ratio."
5. A significant difference was designated as one which would not be expected to occur by chance more than five times in one hundred comparable samples.
6. The procedures used here did not provide control over intergroup differences in the pattern of original registrations. The inadequacy is apparent at this point for 61% of the sons of farmers were enrolled in the College of Agriculture where it was subsequently noted that their persistence was unusually high.
7. E. g., curricula offered by the Colleges of Agriculture, Commerce and Engineering as contrasted with the General Curricula in the College of Liberal Arts and Sciences.
8. Each time this test was made, aptitude scores were divided into upper and lower one-half. The hypothesis of no difference was accepted when the chi square test was not significant at the 5% point.

REFERENCES

1. McNeely, John H. College Student Mortality. U. S. Department of Education Bulletin No. 11, 1937. Washington: U. S. Government Printing Office, 1938.
2. Slater, Marlowe. "Perception: A Context for the Consideration of Persistence and Attrition among College Men," Personnel and Guidance Journal, XXXV, No. 7 (March, 1957), 435-440.
3. United States Employment Service. Dictionary of Occupational Titles, Part II, Titles and Codes. Washington, D. C.: United States Government Printing Office, 1939.
4. Iffert, Robert E., Retention and Withdrawal of College Students, United States Office of Education Bulletin No. 1, 1958, 177 pp.
5. University of Illinois. Undergraduate Study, 1953-1954. University of Illinois Bulletin, L, No. 78. Urbana, Illinois: University of Illinois, July, 1953.
6. Wolfle, Dael. America's Resources of Specialized Talent. New York: Harper and Brothers, 1954.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

THE POSITION OF CERTAIN VARIABLES IN THE PREDICTION OF PROBLEM-SOLVING IN ARITHMETIC

CLINTON I. CHASE
Idaho State College

THE VERBAL problem has long held an important place among the methods of instruction in elementary school arithmetic. Nevertheless, there is a general lack of agreement among writers as to just which of the many variables studied significantly influence one's ability to solve verbal problems.

Some investigators have studied a selected skill (2, 12) or group of skills (5, 8, 13) in their relation to the ability to find solutions to verbal problems in arithmetic. Other studies (10, 14) have identified some causes of difficulty in solving problems, while still others (1, 7) have demonstrated the advantage of a given method of problem analysis as an aid to problem-solving. A few studies (8, 16) have described differences between good and poor problem solvers, and some investigators (3, 4, 11, 17) have illustrated the improvement in problem-solving ability which resulted from training in a given academic skill. Throughout all of these studies, however, no group of variables can clearly be identified as being consistently related to problem-solving ability.

Although it is evident that there is some relationship between a number of abilities and success in the solution of verbal problems in arithmetic, the nature of this relationship is not clear. It is possible that a given skill may be directly related to success in solving problems, or it may evidence only a secondary relationship with a measure of problem-solving ability by way of a closer association with a second variable, which in turn may hold a direct relationship with skill in solving problems. Therefore, the following study is designed to identify those abilities which are primarily related to success in solving problems, and also the variables which evidence only a secondary association with problem-solving ability in that they are more effective predictors of one of the variables which demonstrate that primary relationship.

The Problem

Specifically stated, the problem of the present study is: What are the skills and intellectual factors which are primarily related to the ability to solve verbal problems in arithmetic, and what are the respective contributions of these variables to the multiple prediction of problem-solving ability; and what variables are secondarily related to problem-solving ability in arithmetic, and what are their individual contributions to the multiple prediction of the respective primary variables?

The terms, primary and secondary variables, appearing in the above statement may need definition at this point. In the multiple correlational analysis of a statistical problem, a maximum prediction of the criterion may usually be obtained while utilizing a parsimonious number of variables. In the present study the variables which entered into this parsimonious number, and which in combination provided the best prediction of problem-solving scores, were designated as primary variables.

The secondary variables were those which, in a combination composed of a parsimonious number of variables, provided a maximum prediction of a primary variable.

Procedure

The data for the study were collected from 119 sixth-grade children from two elementary schools in Oakland, California. Their ages ranged from ten years and three months to eleven years and eleven months, the mean age being eleven years and five months. There were 58 girls and 61 boys in the group.

Paper-and-pencil tests were used for the collection of data. Accordingly, the criterion test was the problems section of the Iowa Every Pupil Test of Arithmetic, Form N. There were fifteen independent variables which were measured by the following tests^2:

1. PMA Verbal
2. PMA Reasoning
3. PMA Space
4. PMA Perceptual Speed
5. PMA Number
6. Computation: Iowa Every-Pupil Test of Arith-

---

* Footnotes will be found at the end of the article.

metic

7. Fundamental Knowledge: Iowa Every-Pupil Test of Arithmetic

8. Arithmetic Vocabulary

9. Problem Analysis A: What does the problem ask you to do?

10. Problem Analysis B: What Arithmetic process should be used?

11. Problem Analysis C: Estimating Answers

12. Gates Basic Reading A: Reading for General Significance

13. Gates Basic Reading B: Predicting Outcomes

14. Gates Basic Reading C: Understanding Directions

15. Gates Basic Reading D: Reading to Note Details

All tests used were administered by the author of the present study, with the exception of the reading tests. These were administered by the classroom teachers. To provide maximum familiarity with the testing situation, students took the examinations in their own classrooms, sitting in their customary seats. A period of six weeks was required to carry out the collection of data.

A multiple correlation technique was used in the analysis of the test results, the general design being a variation of the procedure used previously by Holmes (9) in his analysis of skills which lay in the substrata of reading ability. All independent variables were first combined into a multiple correlation, using problem-solving scores as the criterion. Then that independent variable which contributed least to the variance in the criterion was eliminated, a new R computed with the reduced number of variables, and the significance of the differences be tween the new R and the previous R was computed.

This procedure for eliminating variables and achieving a parsimonious number of predictors was repeated until only those variables remained, the elimination of any one of which would significantly reduce the size of R at the five percent level of confidence. When problem-solving scores were the criterion, those variables which were retained as predictors became the primary variables. Each of these primary variables was, in turn, established as a criterion, and a parsimonious number of predictors for each of these variables was selected by the procedure described above. The IBM 701 computer was used for the computations.

Results

By means of the procedure described above, the following results were obtained. Only three of the fifteen independent variables were identified as primarily related to the ability to solve verbal problems in arithmetic. These tests and the variance in the criterion, both direct and shared (6), associated with each variable are given in Table I. Here the test of computation is shown to be associated with a total of 32.33% of the variance in scores in problem-solving ability, Reading to Note Details is associated with 15.93% of the variance, and 13.68% of the variance may be attributed to Fundamental Knowledge in Arithmetic. There remained 38.06 % of the variance in problem-solving scores which was not associated with any of the three primary variables.

When the above primary variables were separately analyzed by the prescribed procedure, secondary variables were identified. The secondary variables associated with Computation are reported in Table II. Fundamental Knowledge accounted for a total of 23.81% of the variance in Computation, while 19.81% of the variance was attributed to PMA - Number, and 11.85% to PMA - Verbal, leaving a total of 44.53% of the variance in Computation not associated with any of the secondary variables.

The next primary variable--Reading to Note Details--was then analyzed, and the results shown in Table III were obtained. Here Reading for General Significance is shown to have been associated with a total of 31.15% of the variance in the criterion, Reading to Predict Outcomes was associated with 24.01% of the variance, Reading for Directions with 21.47%, PMA-Verbal with 9.71%, PMA Number with 5.45%, and PMA Reasoning with a minus 6.25%. There remained 14.46 % of the variance in Reading to Note Details which was unaccounted for.

The last of the primary variables, Fundamental Knowledge, produced the results shown in Table IV. Here Computation is revealed to be associated with 22.97% of the variance in Fundamental Knowledge, Arithmetic Vocabulary with 19.69%, and PMA-Verbal with 12.34%. Forty-five percent of the variance in Fundamental Knowledge remained unaccounted for.

A diagrammatic presentation of the relationships reported in the four tables is provided by Figure 1. From this figure the variables associated with problem-solving ability may be conceived of as representing a hierarchy. Certain skills appear to be primarily associated with the ability to solve problems, while other variables are only secondarily related to this ability in that they are effective predictors of one of the primary variables.

From the results of the method of analysis used in this study, it appears that, in the population of which the present subjects are a sample, in presicting a student's ability to solve verbal problems in arithmetic, as sources of basic data one may turn with considerable confidence to that pupil's ability in the mechanics of computation, his comprehension of the principles which underly the number system, and the extent to which he notices specific items of information when he reads.

Instruction in these skills, it may be noted, are part of the curriculum of the elementary school. They are abilities which instruction proposes to advance. The contention that increased emphasis on these skills in elementary school instruction will result in improvement in the ability to solve verbal problems in arithmetic is supported by evidence

TABLE I

THE PRIMARY VARIABLES AND PERCENTAGE OF VARIANCE IN SCORES IN PROBLEM-SOLVING ASSOCIATED WITH EACH

Variable	Variance directly associated with problem solving	Total variance shared with other variables	Total variables associated with problem solving
1. Computation	20.43	11.90	32.33
2. Reading to Note Details	7.08	8.85	15.93
3. Fundamental Knowledge	4.53	9.15	13.68
Totals	32.04	29.90	61.94

TABLE II

SECONDARY VARIABLES ASSOCIATED WITH COMPUTATION AND PERCENTAGES OF VARIANCE IN COMPUTATION ASSOCIATED WITH EACH

Variable	Variance directly associated with Computation	Total variance shared with other variables	Total variables associated with Computation
1. Fundamental Knowledge	14.37	9.44	23.81
2. PMA Number	12.87	6.94	19.81
3. PMA Verbal	4.51	7.34	11.85
Totals	31.75	23.72	55.47

TABLE III

SECONDARY VARIABLES ASSOCIATED WITH READING TO NOTE DETAILS WITH PERCENTAGE OF VARIANCE ASSOCIATED WITH EACH

Variable	Variance directly associated with Reading Details	Variance shared with other variables	Total variables associated with Reading Details
1. Reading for General Significance	13.57	17.58	31.15
2. Reading for Directions	6.30	15.17	21.47
3. Reading to Predict Outcomes	8.76	16.25	24.01
4. PMA Verbal	1.86	7.85	9.71
5. PMA Number	1.59	3.86	5.45
6. PMA Reasoning	2.40	-8.65	-6.25
Totals	34.48	51.06	85.54

TABLE IV

SECONDARY VARIABLES ASSOCIATED WITH FUNDAMENTAL KNOWLEDGE AND PERCENTAGE OF VARIANCE ASSOCIATED WITH EACH

Variables	Variance directly associated with Fundamental Knowledge	Variance shared with other variables	Total variables associated with Fundamental Knowledge
1. Computation	13.25	9.72	22.97
2. Arithmetic Vocabulary	9.99	9.70	19.69
3. PMA Verbal	4.08	8.26	12.34
Totals	27.32	27.68	55.00

FIGURE 1.

A DIAGRAMMATIC REPRESENTATION OF THE ANALYSIS OF THE DATA SHOWING PRIMARY AND SECONDARY VARIABLES AND TOTAL PERCENTAGE OF VARIANCE IN THEIR CRITERIA ASSOCIATED WITH EACH VARIABLE

	Primary Variables	Secondary Variables
Problem Solving	Computation 32%	Fundamental Knowledge 24%
		PMA Number 20%
		PMA Verbal 12%
		Variance unaccounted for 44%
	Reading to Note Details 16%	Reading for General Significance 31%
		Reading for Directions 21%
		Reading to Predict Outcomes 24%
		PMA Verbal 10%
		PMA Number 5%
		PMA Reasoning -6%
		Variance unaccounted for 15%
	Fundamental Knowledge 14%	Computation 23%
		Arithmetic Vocabulary 21%
		PMA Verbal 12%
		Variance unaccounted for 45%
	Variance unaccounted for 38%	

provided by other investigators (10, 14, 15,17). Since the secondary variables reported in this study demonstrated a significant association with the primary variables, these secondary variables may be investigated as an approach to methods of instruction which may promote competence in the primary variables, and, hence, competence in problem-solving ability itself.

Conclusion

From the results obtained by the analytical techniques used in this study, several conslusions seem to follow.

1. The number of skills primarily associated with the ability to solve verbal problems in arithmetic are relatively few; however, a number of other variables hold a secondary relationship to problem-solving ability in that they are effective predictors of the variables which hold that primary association with the ability to solve problems.

2. From the fifteen independent variables included in this study, a combination of three skills--the ability to compute, to note details in reading, and knowledge of the fundamental concepts in arithmetic--predict problem-solving ability with essentially the same effectiveness as a combination of all fifteen variables.

3. The variables which seem to be secondarily related to problem-solving ability in that they are closely associated with the primary variables are:
   a) intellectual factors, especially the verbal and number factors;
   b) knowledge of the generalizations which underly the number system;
   c) ability to apply reading skills to a variety of purposes.

4. The variables which are primarily associated with performance in solving problems, and many of the abilities which are associated with these variables, are skills which are taught in the elementary school. Related studies support the contention that greater emphasis on these skills may result in a significant increase in the ability to solve verbal problems in arithmetic at the elementary school level.

FOOTNOTES

1. The data for this study were taken from a Ph. D. thesis submitted to the Graduate Division of the University of California, Berkeley, California.
2. The SRA edition of the Primary Mental Abilities was used in this study. Tests 8, 9, 10 and 11 were constructed by the author of the present investigation.

REFERENCES

1. Clark, John R. and Vincent, Leona E. "A Comparison of Two Methods of Arithmetic Problem Analysis," Mathematics Teacher, 18: 226-35, April, 1925.
2. Coffing, Esther A., "The Relationship Between Silent Reading Ability and Arithmetic Ability," School Science and Mathematics, 41: 10-14, January, 1941.
3. Dickey, John W. "The Value of Estimating Answers to Arithmetic Problems and Examples," Elementary School Journal, 35:24-31, September, 1934.
4. Dresher, Richard. "Training in Mathematics Vocabulary," Educational Research Bulletin, 13:201-204, November, 1934.
5. Engelhart, Max D. "The Relative Contribution of Certain Factors to Individual Differences in Arithmetic Problem Solving," Journal of Experimental Education, 1:19-27, September, 1932.
6. Engelhart, Max D., "The Technique of Path Coefficients," Psychometrika, 1: 287-293, December, 1936.
7. Hanna, Paul R. "Methods of Arithmetic Problem Solving," Mathematics Teacher, 23: 442-50, November, 1930.
8. Hansen, Carl W. "Factors Associated with Successful Achievement in Problem Solving in Sixth Grade Arithmetic," Journal of Educational Research, 38: 111-118, October, 1944.
9. Holmes, Jack A. "Factors Underlying Major Reading Disabilities at the College Level" Genetic Psychology Monographs, 49:3-95, January, 1954.
10. John, Lenore. "Difficulties in Solving Problems in Arithmetic," Elementary School Journal, 31: 202-215. November, 1930.
11. Johnson, Harry C. "The Effect of Instruction in Mathematical Vocabulary upon Problem Solving in Arithmetic," Journal of Educational Research, 38:97-110. October, 1944.
12. Lutes, Olin S. "An Evaluation of Three Techniques for Improving Ability to Solve Arithmetic Problems," University of Iowa Monographs on Education No. 6, Iowa City: University of Iowa, June, 1926, 41 pp.
13. Monroe, W. S. and Engelhart, Max D. "The Effectiveness of Systematic Instruction in Reading Verbal Problems in Arithmetic," Elementary School Journal, 33:377-381, January, 1933.
14. Morton, R. L. "An Analysis of Errors in the Solution of Arithmetic Problems," Educational Research Bulletin, 4:187-90, April, 1925.
15. Thiele, C. L. "The Mathematical Viewpoint Applied to the Teaching of Elementary School Arithmetic," The Tenth Yearbook, National Council of the Teachers of Mathematics, New York: Bureau of Publications, Teachers College, Columbia University, 1935, pp. 212-32.
16. Treacy, John P. "The Relationship of Reading to the Ability to Solve Arithmetic Problems," Journal of Educational Research, 38: 86-94, October, 1944.
17. Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," Elementary School Journal, 22:380-86, January, 1922.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

THE QUICK WORD TEST*

EDGAR F. BORGATTA, Cornell University
RAYMOND J. CORSINI, Daniel D. Howard Associates

"INTELLIGENCE" TESTS of the group-verbal type measure capacity in culturally trained skills, or basic mental ability plus experience and training. Practically, they serve as good sources of information about individuals' academic potential. Regardless of arguments that they do not separate out basic factors of mental ability, they do serve as useful indicators of relative ability to do particular kinds of symbolic work.

Development of the Quick Word Test

The Quick Word Test was planned to meet a need for a rapid economical objective instrument, easily administered and quickly scored, for academic, industrial, and institutional populations. In view of the common necessity in testing programs to include other information such as aptitudes, interests, abilities, and personality data, the less time spent measuring general mental ability, the more time can be spent obtaining data in these other areas.

The Quick Word Test is based on the general agreement of psychometricians that the best single indicator of mental ability is understanding of the meaning of words. Words of any length can have diversity of meaning and can be of any degree of difficulty, and on the basis of this argument a standard college type of dictionary was used as a source of four- and five-letter words, obtaining a preliminary panel of approximately 1,700 words and definitions. Peculiar, unusual, and restricted words were screened out of the list, and this reduced the number of words to 1,200 that were then tested on samples of fourth and eighth term high school students, and superior college graduates. The lower grades were used to determine the relative difficulty of the combinations, and the high-criterion group to determine the validity of discrimination of the words. After testing, 700 words were retained.

The word combinations in the Quick Word Test have the following specifications: (a) Eighth term high school students did better than fourth term students. (b) The top quarter of fourth and eighth term high school students did better than the respective bottom quarters, and the college group did better than the top quarter of the high school students. (c) The degree of response to the "correct" answer by the college group substantially exceeded chance. (d) Each of the alternate answers produced for the lower quarter of the fourth term high school students, at least a minimum of responses.

The words were assembled in "blocks" of five in such manner to make each block approximately equal in overall difficulty. Generally, the first word in each block is of minimum difficulty. The first block is no more and no less difficult than the last block. This unusual form has several advantages. First, it keeps the interest of the subject to the very end. Second, it provides a convenient form for scoring since no matter how much an individual does, his score is the percentage of words gotten correctly. If a subject does only one-half of the test, his percentage can be obtained just as conveniently as if he did the entire test. If a subject is given several tests, his percentage correct for each of the parallel forms, or for all forms together, can be obtained. Naturally, the more items a subject does, the greater the reliability of his score.

Properties of the Test

Experience with the test indicates that the objectives are generally achieved. To indicate both objectives and experience, the following list of properties describes the test:

1. The QWT is self-administering. High school students and adults have no difficulty in grasping structions immediately.
2. The test is contained entirely on a single page, and the format is both attractive and non-threatening.
3. Since the QWT is a power test and is not timed, some of the pressure is taken off respondents.
4. Most high school and adult groups complete the test in 8 to 10 minutes.
5. The test is sufficiently reliable. There are

---

* Footnotes will be found at the end of this article.

100 items on the single page form, more than occur in many of the common and more elaborate general purpose tests. Using only one column of the test (50 items) may be sufficient in many situations, and completion time for half the test is 4 to 5 minutes.

6. The test has six independent general forms. In addition, a High Difficulty Form is available for testing college and post graduate groups. Two overlapping junior forms are available for testing retarded groups.

7. The test is valid. It correlates with well-known tests as well as they correlate with each other, and it correlates with some of them at their own reliability levels.

8. The test is economical because it is easy to administer, occupies only a single page, and is rapidly scored with a simple stencil.

At the time of writing, sets of data for two general well-distributed population samples have been analyzed, one from the New Jersey Reformatory at Bordentown, N. J., and one from Central Prison, Raleigh, N. C.

In the Bordentown sample, the Quick Word Test correlated with the Stanford Achievement Test r = .85, the same level at which the Kuhlmann-Anderson correlated with the Stanford for the same sample (See Table I). The QWT correlated with the Kuhlmann-Anderson with r = .78. In the Raleigh sample, the QWT correlated with the Beta I. Q. r = .44, which was the same level at which the Wide Range Achievement and the WRA arithmetic subscore correlated with the Beta I. Q. for the same sample (See Table II). The QWT, it should be emphasized, was as efficient for this purpose when only one-half the test was used (only 50 words). The QWT correlated with the WRA with r = .80. While the QWT correlated higher with the spelling subtest of the WRA than with the arithmetic, it correlated with the Beta I. Q. at the level of the arithmetic subtest. In both these general samples, the cross-test reliability of the QWT was well over r = .90, and the reliability for half the test was also in this neighborhood.

The QWT was administered to samples of female students in the second year of high school. For this homogeneous group, as indicated in Table IV, the correlation with the California Short-Form Test ranged from r = .53 to .80, with nonoverlapping subsamples of varying size. Considering the restricted variance, these correlations are substantial. A reliability check was carried out with a sample of college students, primarily juniors in a school of education evening class, but with a few seniors in the sample as well. For this restricted group, the cross-test correlations are presented in Table III. Correlations were computed for Forms A, B, C, and the High Difficulty Form.

It should be noted that the correlation from the general tests to the High Difficulty Form are somewhat lower than those among the general tests. The utility of the High Difficulty Form is not yet fully explored, but it is designed for screening persons at higher levels and measuring the qualities of maturity associated with continued intellectual interests. In Table III, it is seen that the High Difficulty Form is somewhat higher correlated with age than the other forms. Emphasis has been on general concepts in the selection of words for the test, rather than specific, local, or idiosyncratic types of concepts. It may turn out that, in discerning meaning, distinguishing a definition or synonym from an association requires somewhat different kinds of intellectual capacities as the difficulty of the concept increases[2]. Preliminary experience and current experimental use of the High Difficulty Form suggests it may prove to have qualities that will make it particularly useful for the study of brilliant and high achievement groups.

Summary

The QWT has been designed to assess the general capacities commonly included when one thinks of a measure of intelligence. Our objectives have been to maximize simplicity, speed, and economy. Both from the point of view of the tester and the subject, these objectives seem to be satisfied. The test appears to have exceptional reliability. It is obvious the QWT taps the same content as other similar tests commonly encompass. From the point of even greater economy of resources, in use with general (well-distributed) populations half the test should prove to be satisfactory, thus reducing the general administration of the test to 4 or 5 minutes. Coupled with the convenience of being an untimed test, these are remarkable characteristics for such a test.

FOOTNOTES

1. Many persons have contributed to the development of the QWT. As a project, it would have been virtually impossible without the contribution made at all stages by Mary Weed McGurk. At various times, Jonathan Robbin, John Stimson, and Robert F. Guerrin have given technical assistance. Dr. Walter Wolf and Milton Novak, of Bryant High School, Queens, New York City, were particularly helpful in collecting the initial criterion data; Dr. Elliott H. Stofflet, New Jersey Reformatory, Bordentown, N. J., James H. Panton, Central Prison, Raleigh, N. C., and Lenore R. Miller, High School of Fashion Industries, New York City, assisted generously in providing data reported here.

2. In distinguishing the correct meaning in these words, the task requires more than mere recognition. In fact, since the words used as "misleaders" are specifically chosen for their value as possible associations, something more than familiarity with the language appears to operate. Interviews with respondents indicate that those who do best on the High Difficulty

TABLE I

BORDENTOWN SAMPLE

Test	Mean	Sigma	Form A	Form B	Kuhlmann-Anderson	Stanford
QWT Form A	53.3	21.29		.95	.78	.84
QWT Form B	53.0	21.27			.77	.86
Kuhlmann-Anderson MA	152.8	24.74				.85
Stanford Achievement Grade Equivalents	7.4	2.18				

Note: Characteristics of the total (male) population are:
50% Average or better with few superior, 30% dull normal, 20% borderline and moron;
Age range: 16-30; 63% White, 37% Negro;
A representative sample of 95 cases was selected for use.

TABLE II

RALEIGH SAMPLE (130 MALES)

Variable	Mean	S.D.	Beta I.Q.	Test Age	Spelling	Arithmetic	Form C		Form D	
							Left Half	Total	Left Half	Total
Age	29.9	9.7	02	03	04	-14	11	16	16	19
Revised Beta I.Q.	99.8	10.8		44	34	46	44	44	45	43
WRA--Test Age	12.4	2.8			94	79	80	81	77	78
Spelling	7.2	2.9				68	76	79	74	76
Arithmetic	5.6	2.3					54	54	52	50
QWT Left Half (Form C)	57.0	24.1						95	85	87
Total	52.0	20.2							91	93
QWT Left Half (Form D)	51.8	21.0								98
Total	52.4	20.4								

Note: Decimal points omitted.

TABLE IV

ENTERING SAMPLE, VOCATIONAL HIGH SCHOOL (FEMALES)

Form	1956 Data						1958 Data					
			QWT		California MA				QWT		California MA	
	r	N	Mean	Sigma	Mean	Sigma	r	N	Mean	Sigma	Mean	Sigma
A	.76	17	42	15	154	17	.61	64	39	15	152	18
B	.53	19	36	10	151	18	.72	61	41	16	155	19
C	.72	30	38	13	154	14	.70	52	42	17	156	23
D	.71	75	41	14	152	18	.72	60	42	17	154	21
E	.66	79	37	13	154	19	.77	49	40	15	56	20
F	.63	72	39	12	156	16	.80	49	38	14	154	22

Form recognize the distinction between a definition and an association most clearly, and they operate on the principle of elimination of associations in finding the proper word when it is not known. In other words, aspects of familiarity may be used negatively in the logic that identifies the correct meaning. An interesting characteristic of responses to a number of words on the High Difficulty Form is that persons who do poorly on the test do not choose answers according to "chance," but by association and, thus, the base of "no knowledge" about these words falls closer to zero than to the 25 percent that would be expected by chance.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

DETERMINING INDIVIDUAL NEEDS OF ELEMENTARY TEACHERS AS A BASIS FOR AN ORIENTATION PROGRAM*

ANNE PHILLIPS McCREARY
Cornell University, Ithaca, New York

THE EFFECTIVE induction of teachers with varying background and training into school situations which likewise differ widely has long been a concern of administrators and boards of education. This problem has become more acute in recent years due to teacher turnover, drop-outs, and shortages caused by the increasing size of schools. The administrator must present and interpret policies and procedures to the new teacher and then assist him in adjusting to them. His goal is the development of efficient, confident, happy teachers who will not only become more effective but who will wish to remain on the faculty. A strong orientation program can lead to this goal.

A survey of literature shows that there is a need for better orientation. Eye, in The New Teacher Comes to School (1), lists the many adjustments which the new teacher is expected to make. While many administrators plan orientation programs in an attempt to make this adjustment easier, a study in California by Dutton (2) and one in Indianapolis by Tower (3) seem to indicate that orientation programs now in practice do not always meet the needs of beginning teachers. The various problem areas identified most frequently by research can be organized into these broad categories: discipline, organization, philosophy, and personality. No report has been found of attempts to anticipate problems which the individual beginner is most likely to encounter. Rather research findings have been used in evaluating present procedures and in planning for a more effective program but still on a mass basis with no emphasis on individualization.

The Problem

This study was an endeavor to develop a method for diagnosing individual strengths and weaknesses in candidates for elementary teaching positions as the initial step in planning a more highly individualized orientation program. The assumption was made that the administrator could give more immediate and more effective assistance to the new comer if he had information regarding individual strengths and weaknesses. This information could be obtained during the initial interview and used in planning a more highly individualized orientation program which could lead to more rapid adjustment for the beginning teacher. The purpose of this study was:

1. To develop a technique for securing information from job candidates which would indicate probable strengths or weaknesses in the four areas--personality, philosophy, organization and discipline.
2. To develop an instrument for use in evaluating and recording this information.
3. To analyze the data in order to determine: a. if such a method was feasible and did yield useful information; b. if the four areas designated as problem areas by research findings were revealed as significant problem areas in the sample selected for study; c. if administrators felt that information gathered by such a technique could be used in planning for individual orientation.

Subjects

Administrators or supervisors from twelve school systems were asked to participate in this study. These schools were all from one state but both urban and rural schools were included. All of the rural schools were central school systems. (A central school system is formed when several schools or school districts within a designated geographical area unite under one administration). The forty-three candidates for elementary teaching positions who were interviewed by participating administrators during the month of March were the source of the data.

Procedure

The assumption was made that administrators do use some criteria or basis for evaluating candidates during the initial interview. From this source might come valuable suggestions which had not been previously reported in the literature. Therefore, the twelve administrators were asked to send either a copy of the check list now in use during interviews with job candidates or a list of criteria used in selecting candidates. A diagnostic-evaluative check list was prepared based on this information and on research findings. See Table I. This check list contained seventeen sub-categories arranged under four main headings. Under each sub-category, ratings could be made on a three-point scale or inability to arrive at a rating could be indicated.

---

* This report is part of a dissertation submitted to the Graduate School of Cornell University, 1959.

Administrators were provided with copies of the check list and also with a series of questions which had been designed to elicit responses in the four specified problem areas. It was suggested that the administrator, in conjunction with his initial interview with job candidates, use a semi-structured interview technique. The suggested questions could be omitted, substitutions made, or order varied. After conducting the interview, the administrator was asked to make an evaluation and then to use the diagnostic-evaluative check list, rating each candidate as strong, average, or weak in each sub-category.

In order to ascertain the administrator's opinion regarding the adequacy of the instruments and the feasibility of using this information to plan for individual orientation, a sheet was prepared for his use. On this sheet provision was made for the recording of difficulties encountered in using the technique and instruments, suggested changes and improvements, opinions regarding the adequacy of the instruments, and the usefulness to the administrator for orientation.

Results

One of the twelve administrators who were asked to participate in the study made no response. The two who supervised large city systems indicated that they did not have sufficient time to use this instrument. Two of the administrators in central schools interviewed no candidates during the specified period and thus were unable to contribute data. However, the sample retained its original character as it included the principal in a city school and administrators in central schools.

Table I shows the overall picture when results were totaled and tabulated. Administrators were unable to make judgments on only thirty-six of the seven hundred and thirty-one items. These cases which were checked 'unable to determine' occurred most frequently in attempting to evaluate potential for self-discipline, self-evaluation, and flexibility. Discrimination was possible in evaluating candidates and rating them as strong, average, or weak.

Table II shows the distribution of items judged as weak. Among the four problem areas the distribution was fairly even with the highest percentage of weaknesses indicated in the area of philosophy.

Table III lists the information submitted by administrators on the evaluation sheets. Examination of the opinions from this small sample indicates that the research instruments were considered usable, successful, and valuable in the capacity for which they were designed.

Conclusions and Discussion

From the examination and analysis of the data gathered from this study, certain conclusions may be drawn. These conclusions are based on the information obtained from subjects in a small regional sample. From this rather narrow basis, certain broad generalizations may be made. Implications for further study are also evident.

1. Response to the research instruments indicate that the proposed procedure is most useful to administrators in central schools or to building principals in city schools.

2. In discussing the adequacy of the research instruments several points should be clarified. First, the author, in developing the technique and designing the research instruments, was planning a basic tool for use by administrators during interviews with job candidates. This basic tool could and should be modified or adapted to meet the specific situation. Second, the adequacy of the instruments here refers to adequacy in obtaining and recording information regarding probable strengths and weaknesses of job candidates in four selected problem areas only. Third, adequacy of the instruments can not be determined absolutely. Influencing the judgment of each item on the check list are various factors related to the administrator, the candidate, and the situation. Analysis of the results indicates that the research instruments were practical, usable, and adequate as basic tools. Adaptation would be necessary for use in varied school situations. Information gained through the use of these instruments could be used to plan for individual orientation.

3. The administrators were 'unable to determine' most often in the area of philosophy. In order to interpret and to make evaluations in an area which is as important and as difficult to define as philosophy, one must himself have arrived at a clearly defined statement in that area. Since information about self-evaluation and self-discipline was difficult to obtain, it may be that these sub-categories do not belong on such a check list.

4. The evaluation sheet was used by four of the administrators who tested the proposed procedure. Use of this sheet had not been emphasized in the instructions for use of materials. In the initial stages of planning, this facet was seen as a minor sub-area where information might be obtained. As the study progressed, it was evident that this instrument could be very important in establishing the possible value of the technique as an aid to individual orientation.

LIBRARY
1.8.70
3002

TABLE I

SUMMARY OF DIAGNOSTIC-EVALUATIVE CHECK LIST DATA

Category	Strong	Average	Weak	Unable to Determine	No Response	Doesn't Apply
I Personality						
1. good grooming	32	9	2	0	0	0
2. poise and maturity	31	10	2	0	0	0
3. Positive outlook and high morale	20	20	1	2	0	0
4. liking for people, children and teaching	27	15	0	1	0	0
Totals (172)	110	54	5	3	0	0
II Philosophy						
1. adaptability	19	19	1	3	1	0
2. self-confident but willing to accept help	23	14	4	4	2	0
3. ability for self-evaluation, a realistic point of view	18	17	1	6	1	0
4. a questioning but not hypercritical attitude	16	19	7	1	0	0
5. professional and ethical attitude	16	19	7	1	0	0
Totals (215)	99	84	14	15	3	0
III Organization						
1. recognizes the importance of preparation	30	9	0	3	1	0
2. considers grouping within the room necessary	27	12	2	1	0	1
3. is prepared to accept individual differences	27	15	0	0	0	1
4. shows potential for democratic, flexible operation	18	16	2	6	1	0
Totals (172)	102	52	4	10	2	2
IV Discipline						
1. Recognizes that discipline is necessary for a good teaching-learning situation	35	8	0	0	0	0
2. good discipline is fair, consistent, and suggests a constructive alternative	30	10	2	1	0	0
3. discipline is the teacher's responsibility	30	12	0	0	1	0
4. good discipline involves self-discipline	20	14	2	7	0	0
Totals (172)	115	44	4	8	1	0
Aggregate (731)	426	234	27	36	6	2

Experience of candidates interviewed: Beginner--21 1 year -- 9 2 years-- 5
3 years 1 4 years-- 1 5 years-- 1
6 years-- 2 7 years-- 1 9 years-- 1
11 years-- 1

Total number of candidates--43

TABLE II

DISTRIBUTION OF ITEMS JUDGED AS WEAK

Area	Number of Items	% Of Total Judgments	% Of Weak Judgments
I	5	3	19
II	14	7	52
III	4	3	15
IV	4	2	15
Total	27	4	

TABLE III

INFORMATION FROM EVALUATION SHEETS

Suggestions listed on the sheet	Responses made by subjects			
	No. 3	No. 4	No. 5	No. 12
1. Suggested changes and improvements	Candidate can answer as he thinks he should	None	Need facts about family background and future plans	Add section on communication and keeping of records
2. Difficulties	Easy to use	None	Too time consuming with his own method	None
3. Adequacy of questions	Very good	Clear, good	Quite good. Could be more specific	Good
4. Have you gained information re problems?	Yes	Candidates are screened before interview, thus no problems	No, don't hire teachers with problems	Yes
5. Will you use this information in orientation?	Yes	Yes, with all other information	No	Yes
6. Comments	Follow-up would be good	Good if adapted to meet specific school systems	A good tool, will use some of the questions	Very comprehensive with a few additions

REFERENCES

1. Eye, Glen G. and Lane, Willard. The New Teacher Comes to School, New York: Harper and Brothers, 1956.
2. Dutton, W. H. The Orientation of First Year Teachers to Classroom Service. California Teachers Association. Research Department. (Bulletin No. 52). November, 1952.
3. Tower, Melvin M., "A Study of Orientation and In-Service Education Practices in the Indianapolis Public Schools," Education Administration and Supervision. April, 1956, pp. 219-29.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

A STUDY OF ACHIEVEMENT IN TWO TYPES OF PUBLIC SECONDARY SCHOOLS

JOHN W. FRENCH
Educational Testing Service, Princeton, New Jersey

THE DEMOCRATIC ideals of this country favor a free secondary school education for all students. While, according to this ideal, all educational opportunities should be equal and affluent, the character of geographical regions and the willingness or unwillingness of people to support public education have given rise to differences among schools that may be associated with differences in the quality of education. The educational opportunities offered to students by various kinds of public secondary schools have been investigated by Dr. James B. Conant and the staff of "A Study of the American High School," a project supported by the Carnegie Corporation of New York. To supplement data gathered during many visits to schools, it was desirable to secure evidence based on objective tests as to whether at least some comprehensive high schools do as good a job of teaching as is done by college-oriented schools with a national reputation.

The educational opportunities being considered here are those affecting subject-matter achievement. For purposes of this study, good achievement will be defined as that associated with achievement test scores that are higher than what might be expected on the basis of verbal and quantitative aptitude test scores.

The Testing Program

The aptitude test used in this study was one known to measure general scholastic ability, the Cooperative School and College Ability Test, college level. Form IB of this test was administered, because some of the schools had already given Form IA to some of their students. The verbal score of the SCAT is derived from sentence-completion tasks and vocabulary items. The Quantitative score is derived from numerical computations and mathematics reasoning problems.

The achievement tests administered to measure teaching effectiveness were the Cooperative Plane Trigonometry Test and a past form of the College Entrance Examination Board Physics Achievement Test. Both of these tests represent the thinking of committees composed of high school and college teachers experienced in the construction of objective tests. Form U of the Trigonometry Test was used, because at some schools the students who had taken trigonometry during the first semester of the school year had taken Form Y. It was safely assumed that none of the students had seen the Physics Test, since that is a "secure" test which had not been administered to high school students for several years.

It would, of course, be desirable to study achievement in subjects other than trigonometry and physics, but there are reasons why doing this with the method described here would be either impossible or highly misleading. Social studies or American history achievement were not included because test scores would be affected by the great variability in emphasis of courses in these subjects and by the practice of offering U. S. History sometimes in one and sometimes in two years. Similarly, in the case of foreign languages, there is tremendous variation in number of courses taken, differences in teaching methods (aural or grammatical), and differences in foreign language exposure in the home or while traveling abroad. There seemed little hope that a language achievement examination would reliably measure the quality of teaching in the school rather than factors such as grade placement, amount of travel, or home environment.

It would also be desirable to include tests of proficiency in reading and writing, but the method used here did not seem appropriate for detecting differences in the quality of instruction in those skills. Reading is so highly related to the verbal score of the SCAT that a school which was successful in teaching reading would also raise the scores of its students on SCAT. As a result, the reading ability of its students would not look high in comparison to SCAT scores, and the good teaching might go undetected. To a considerable extent the same thing can be said about detecting differences in writing instruction. In addition, the unreliability of scoring samples of writing and the high cost of this process made it seem particularly unwise to undertake an experiment that might imply the existence of no differences when in fact moderately large differences existed.

Administration of the Tests

The Trigonometry, Physics, and SCAT were administered at 17 public high schools during the last week of April or the first week in May, 1958. All students who were taking or had taken trigonometry during the 1957-1958 school year took the trigonometry test; all students taking physics took the physics test; and all students taking either or both of these took the SCAT.

The schools were asked to avoid informing the students in advance of the nature of the tests and to refrain from any kind of special preparation. They were asked to schedule the tests in such a way that the students could not obtain advance information from each other. There is every reason to believe that these requests were carried out conscientiously.

The Schools

The schools participating in the study were selected by the staff of "A Study of the American High School." All were public secondary schools. One group of 13 was selected from among schools of good reputation which offered comprehensive programs of study and had fewer than 50 percent of their graduates go to college. A second group of four schools was chosen from among institutions with an outstanding national reputation for college preparatory facilities and a high percentage of graduating students going to college. The first group will be referred to as "comprehensive" schools and numbered 1-13; the second group as "college-oriented" schools and numbered 21-24.

Results

Table I gives the deviations of each school mean on the achievement tests from the means that would be predicted on the basis of the SCAT total score. The computations made use of the obtained correlations between SCAT total score school means and trigonometry means of .62 for boys and .56 for girls. The corresponding figures for physics were .62 for boys and .66 for girls.

The deviations may be further explained by example. For School No. 1, with boys and girls combined, the predicted mean on the trigonometry test was 54; the actual mean was 53. Similarly, the predicted mean on the physics test was 501, while the actual mean was 463. It is presumed that positive deviations are associated with effective instruction, while negative deviations are associated with less effective instruction. Since the trigonometry and physics tests used different score scales, the absolute sizes of deviations for trigonometry cannot be compared with those for physics. Published figures show that for individuals (rather than for school means) the standard deviation for the trigonometry score is 8.7 for eleventh grade students who had taken a course in trigonometry. That for the physics scores is about 100 for candidates using the College Board physics test for admission to college.

The principal question asked in this research is answered by the distribution of the deviations. Students in a number of comprehensive schools are definitely not at a disadvantage compared to students in the college-oriented schools. The overlap in the range of deviations is wide. In fact, one of the comprehensive schools appears to have done a better job in teaching both trigonometry and physics (in terms of the tests used) than any of the four college-oriented schools.

If the deviations in Table I are averaged, it will be found, particularly by weighting schools according to size, that the average deviation for college-oriented schools is higher than that for comprehensive schools both in trigonometry a and physics. However, the differences are small in comparison to the total range of deviations. A formal test of the significance of differences between the comprehensive and college-oriented groups of schools does not seem appropriate, because the selection of schools was not intended to be random or representative of defined populations of schools.

Table I also gives some characteristics of the courses given at each school. Trigonometry is a half-year course given sometimes in the first and sometimes in the second semester as noted in the table. It can be reasoned that students taking it in the first semester might be under a disadvantage on an examination given in May because of the lapse of time. On the other hand, students taking trigonometry in the second semester might be under a disadvantage because instruction was not complete. The table gives the number of school days left in the semester after the day of the test. For physics, Table I notes the existence or non-existence of a laboratory as well as the number of school days left after the day of the test. For both courses, Table I indicates whether the curriculum as outlined by the teacher was the traditional one covered by the test or whether it contained non-traditional material. No strong or consistent relationships between the deviations and any of these characteristics seem to be evident.

The question was also asked whether the comprehensive or college-oriented schools have any differential effect on the achievement of students at different levels of ability. Since the number of girls in the physics and trigonometry classes was small in most of the comprehensive schools, only the results obtained by testing the boys were used in this part of the analysis. In order to compare groups with similar ability to handle quantitative problems, the boys in each school were assigned to three groups: those who answered correctly forty-six out of fifty questions on SCAT-Q, those who answered correctly between forty-one and forty-five questions, and those who answered correctly forty or less.

The test performance of boys who studied phys-

TABLE I

DEVIATIONS OF SCHOOL MEANS ON ACHIEVEMENT TESTS FROM PREDICTED MEANS AND CHARACTERISTICS OF COURSES THAT MIGHT ACCOUNT FOR THE DEVIATIONS

School	Trigonometry					Physics				
	N	Deviation	Semester	Days Left	Traditional	N	Deviation	Laboratory	Days Left	Traditional
1	55	-1	2	16	Yes	105	-38	Yes	17	Yes
2	50	0	1 and 2	10	Yes	101	-39	Yes	10	No[1]
3	29	-2	1	0	Yes	32	4	Yes	16	Yes
4	21	-3	2	23	No	12	15	No	22	Yes
5	35	-4	2	17	No	86	19	No	17	Yes
6	78	2	1	0	Yes	65	16	No	16	Yes
7	26	6	2	23	Yes	37	52	Yes	23	No
8	32	-1	1	0	Yes	75	-42	Yes	23	Yes
9	63	-2	1	0	Yes	124	-27	Yes	21	No
10	88	2	2	21	No	81	21	Yes	20	Yes
11	47	0	1	0	No	54	-19	Yes	20	Yes[1]
12-1[2]	70	-1	1	0	Yes	119	9	Yes	24	Yes
12-2[2]	69	-1	2	24	Yes					
13	44	5	2	15	Yes	37	-15	Yes	14	Yes
21	96	4	1	0	Yes	91	-14	Yes	14	No
22	118	4	1	0	No	203	31	Yes	24	Yes
23	153	2	2	27	No	98	-5	No	27	Yes
24-1[3]	59	-3	2	22	Yes	51	36	Yes	17	Yes
24-2[3]	43	-5	2	22	Yes					

These two schools reported having a substitute teacher in physics during the whole year.

School 12 figures are shown separately for the first and second semester trigonometry students.

School 24 has a senior trigonometry section (24-2) which is more able and has had a more extensive mathematical background than the junior trigonometry section (24-1).

TABLE II

PHYSICS TEST RESULTS FOR BOYS AT THREE LEVELS OF QUANTITATIVE ABILITY

Ability Level	Type of School	SCAT-Q Score		Physics Score	
		Average of School Means	Range of School Means	Average of School Means	Range of School Means
High	College-Oriented	332	331-333	588	564-609
	Comprehensive	331	329-334	567	502-692
Medium	College-Oriented	320	319-320	541	506-598
	Comprehensive	320	318-320	509	459-568
Low	College-Oriented	307	303-309	504	474-535
	Comprehensive	303	294-307	454	417-513

TABLE III

TRIGONOMETRY TEST RESULTS FOR BOYS AT THREE LEVELS OF QUANTITATIVE ABILITY

Ability Level	Type of School	SCAT-Q Score		Trigonometry Score	
		Average of School Means	Range of School Means	Average of School Means	Range of School Means
High	College-Oriented	332	329-334	62.9	55.2-66.9
	Comprehensive	331	326-334	59.5	54.1-66.7
Medium	College-Oriented	320	319-321	56.4	52.7-60.0
	Comprehensive	320	319-321	54.2	47.5-61.2
Low	College-Oriented	306	304-309	52.4	46.8-57.9
	Comprehensive	306	299-313	50.0	44.9-58.7

ics is summarized in Table II. The college-oriented schools had relatively more students in the high ability group and fewer in the low group. However, the mean scores on SCAT-Q for members of each group are quite uniform from one kind of school to the other as shown in the table. Therefore, it is possible to make direct comparisons between the physics test scores. Although the averages of school means in physics show differences at all levels of ability which favor the college-oriented schools, the ranges of school means demonstrate a good deal of overlap for all ability groups between the apparent effectiveness of instruction in the two kinds of schools.

Table III presents data for boys taking trigonometry. The results are similar to those for physics. The differences between average school means in trigonometry again favor the college-oriented schools for all levels of ability, but there is much overlap between the two types of schools.

These so-called high, middle, and low ability groups are quite able students. The publisher's norms show that the average student in the high ability group is in the top one percent of a cross-section of public high school seniors, the average student in the middle group is at the ninety-third percentile, and the average student in the low ability group for comprehensive high schools—the least able subsample of all— is at the eighty-second percentile.

The scores on the Physics Test indicate a high level of competence in the subject. The score 500 is the expected score for a twelfth-grader whose mathematical ability has been estimated to be at approximately the seventy-fifth percentile among the twelfth-graders throughout the nation. Many of the school means referred to in Table II are far above 500.

Discussion

In this study the "teaching effectiveness" of a school is taken to be a function of the mean score on achievement tests with mean aptitude test score held constant. It is recognized, however, that this is an oversimplification. Curriculum emphasis, attitude of parents and fellow students, out-of-school reading, habits and attitudes acquired in earlier schooling, and general educational background can all affect the measure of teaching effectiveness so defined. The conclusions made from this study must, therefore, rest on the assumption that any important effects of these things on the scores of individuals cancel out when the school means are considered.

In addition, it is necessary to assume that the achievement tests reflect all of the essential achievements brought about by the trigonometry and physics courses in any of the participating schools. That is, a school may give something extra and important that is non-traditional to its students, but this will only be detected, in an experiment like this, if the non-traditional materials are included among the test items. This assumption limits the conclusions of the present study. However, since the tests represent some of the best thinking by school and college teachers, it is very useful to know that there is wide overlap between the two types of school with respect to their ability to teach trigonometry and physics as it is represented by these tests.

Insofar as the evidence provided by this study is concerned, there is marked overlap between the two groups of high schools with respect to teaching effectiveness. Such small differences as were observed between the means for the two types of schools were judged to be unimportant in the interpretation of the results. A significance test of differences found between types of school was not reported because the study was not designed for this purpose. The samples of schools cannot be considered random or representative samples of defined populations of schools. Any differences found must be interpreted in terms of the particular schools selected.

While this limitation renders it impossible to make any definitive statements about comprehensive schools in general, an answer can be made to the question which the study was designed to answer. The staff of "A Study of the American High School" wanted to find out whether at least some comprehensive high schools bring about achievement (as represented by the tests) equivalent to that made by the college-oriented schools. The conclusion is clear that many of the comprehensive schools can and do bring about such achievement.

Summary

In connection with their investigation of the problems of American secondary education, Dr. James B. Conant and the staff of "A Study of the American High School" wanted to obtain objective evidence on the question whether there are comprehensive high schools which do as good a job of teaching as is done by highly reputed college-oriented schools. Final achievement in trigonometry and physics courses in 13 comprehensive schools and four college-oriented schools was measured by appropriate Cooperative and College Board tests. Ability was held constant in the analysis by scores on SCAT. A very large overlap in the range of teaching effectiveness, as reflected by these tests, was found for the two types of schools. The same kind of overlap persisted when the data were analyzed separately for students at three different levels of ability.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

THE ADEQUACY AND USEFULNESS OF AN OBJECTIVE LANGUAGE SCALE WHEN ADMINISTERED TO ELEMENTARY SCHOOL CHILDREN*

WILLIAM G. WILLIAMS
Purdue University, Lafayette, Indiana

Purpose

THIS STUDY attempted to appraise the adequacy and usefulness of an objective scale designed to measure language development, when it is administered to elementary school children, ages five through twelve years. This scale is a modification of one developed earlier by Dr. M. J. Mecham.

Hypotheses

1. There is no significant relationship between the language scale and the Metropolitan reading tests.
2. There are no significant differences among the scores obtained by elementary school children as a function of grade level.
3. In speech-delayed children there is no significant relationship between their scores on the language scale and their ratings of language development as evaluated by trained speech clinicians.
4. The language scale used in this study is reliable.
5. There is no significant difference between the language scale scores obtained by the trained clinician and those obtained by a teacher.
6. There is no significant relationship between the results obtained by the teachers' use of the scale and the teachers' subjective evaluation of the children's level of language.

The Subjects

A total of 140 elementary school children were randomly selected from two schools representing middle and extreme ranges of socio-economic families in the Provo, Utah area. This group consisted of five boys and five girls from each grade level in each school, kindergarten through sixth grades.

The stimulus consisted of a language scale listing items of language development at successive age levels.

Procedure

1. Each child was administered the direct-test items in the language scale. Raw scores were used as measures.
2. The remaining language scale items were obtained in parent interview, making a total language score which was used as raw data.
3. Metropolitan reading tests were administered to the children, at each grade level.
4. Two evaluations were made of each child by his teacher: a subjective evaluation of each child's language, and administration of the language scale.

Analysis of the Data

1. Scores on the language scale and Metropolitan results were tested by means of rank-order correlation. A significant relationship was found, except in three grades.
2. The language scale scores were analyzed by simple analysis of variance, F. A significant difference was found in successive grade levels.
3. The scores of speech-delayed children were correlated with the evaluations of speech clinicians. A significant relationship was found between these two measures.
4. The language scale was tested for reliability by use of the split-half technique. A significant correlation of reliability was found.
5. The scores obtained by teachers and clinician, all using the language scale, were tested by t-ratio and rho correlation. A significance of difference was found with t; a significant relationship was found with rho for most grades.
6. The teachers' subjective evaluation and use of the language scale were correlated by using rho and the sign test. No significant relationship was

---

* An abstract of a thesis submitted to the Department of Speech and Hearing Rehabilitation, Brigham Young University, Provo, Utah, in partial fulfillment of the requirements for the Master of Arts degree, August, 1958.

LANGUAGE DEVELOPMENT SCALE USED IN THIS STUDY

0-1 Year Level

S 1. Crows, laughs or smiles.
S 2. Produces consonant sounds reflexively.
S 3. "Talks," imitates sounds.
L 4. Responds to name and "no-no."
L 5. Comprehends "bye-bye" and "patacake."
S 6. Echoes words or sounds (Dada or Mama).

1-2 Year Level

S 7. Expressive vocabulary of at least two words.
L 8. Follows simple instructions.
L 9. Reacts to name--recognizes it when called. *
W 10. Marks with pencil or crayon.
L 11. Recognizes names of at least 6 or more common objects.
S 12. Uses names of familiar objects.
R 13. Picks up books; leafs through them. *
S 14. Talks in short sentences.

2-3 Year Level

S 15. Uses "I", "me", or "you" in speech.
S 16. Expressive vocabulary of at least 25 words in speech.
L 17. Identifies common pictures when they are named.
S 18. Can name most common pictures.
S 19. Says full name when asked.
S 20. Asks for another (more, some more, etc.).
L 21. Recognizes hair, mouth, ears, hands, when they are named.
S 22. Verbalizes toilet needs.
S 23. Uses some plurals.
R 24. Distinguishes between big and little in pictures. *

3-4 Year Level

S 25. Expresses vocally his desire to take turns.
S 26. Names almost all common pictures.
R 27. Identifies action in pictures. *
S 28. Says at least one nursery rhyme or sings nursery song.
S 29. Vocabulary of 50 or more words in his speech.
S 30. Names one color.
R 31. Matches common pictures (identical) on lotto board. **
S 32. Has mastered m, p, b, w, and h sounds in his speech. **

4-5 Year Level

S 33. Names all colors.
S 34. Has mastered t, d, n, j, k, g and ng sounds in his speech. **
L 35. Repeats 4 digits, one second apart. **
R 36. Reads by way of pictures.
S 37. Recites numbers one to ten. *
R 38. Recognizes own name in print. *
S 39. Relates experiences.
L 40. Answers telephone; takes simple message. *
S 41. Names penny, nickle, and dime.

5-6 Year Level

S 42. Can tell a familiar story or fairy tale.
L 43. Listens attentively. *

W 44. Prints first name. *
S 45. Sings songs learned in school. *
L 46. Can repeat a 12 syllable sentence. **
W 47. Draws recognizable forms with pencil or crayon.
S 48. Has mastered v, sh, zh, ch, th (as in there) and l sounds in speech.**
L 49. Can tell difference in musical tempo. *
S 50. Recites numbers to thirty.
W 51. Prints simple words (copying). *

6-7 Year Level

W 52. Prints full name alone. *
S 53. Relates fanciful tales.
W 54. Prints simple words (without copy).
W 55. Prints numbers one to fifty. *
R 56. Reads on pre-primer level.
R 57. Reads in average reading group in class. *
S 58. Recites poem or sings song before group.
S 59. Asks meanings of words.

7-8 Year Level

L 60. Can repeat series of five digits, one second apart. **
S 61. Names quarter, half-dollar, and dollar.
S 62. Can name 25-30 separate words one after the other. **
W 63. Prints or writes with a pencil twelve or more words (without copy).
S 64. Has mastered s, z, wh, r, and th (as in thing) in speech. **

8-10 Year Level

R 65. Reads on own initiative.
S 66. Can retell a short story he has read on his own.
L 67. Can repeat a 16 syllable sentence. **
W 68. Writes occasional short letters.

10-12 Year Level

L 69. Can repeat a series of six digits, one second apart. **
L 70. Can repeat a 20 syllable sentence. **
R 71. Enjoys books, magazines, and newspapers.
W 72. Answers ads, purchases by mail.
L 73. Can repeat five one-syllable words. **
W 74. Communicates by letter; serious information.
S 75. Follows current events, discusses them with others.

** Direct test items.
* Items added by this investigator in an attempt to broaden the number of representative areas.

Most of the items were taken from Doll's Vineland Social Maturity Scale, Gesell's Developmental Schedules, Terman and Merrill's revision of the Binet Scale, Poole's Maturation Scale of Articulation, and the works of McCarthy.

Key to Categories

R--Reading
W--Writing
S--Speaking
L--Listening

found with most grades.

Definition of the language scale items and other explanatory data may be obtained by writing the author at Purdue University, Speech and Hearing Clinic.

Conclusions

1. The language scale is adequate and useful for measuring language development in elementary school children.

2. The language scale measures progressive language development as a function of grade level.

3. The language scale has apparent validity in its rating of speech-delayed children.

4. The language scale used in this study is reliable.

5. Teachers apparently may use the language scale adequately without extensive training.

6. The teacher's use of the language scale is a more objective evaluation of language development than is her subjective evaluation.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

COMPARISONS OF GOOD AND POOR READERS IN FOURTH AND EIGHTH GRADES

IRA E. AARON
University of Georgia
Athens, Georgia

CONFLICTING opinions are frequently expressed by teachers concerning the relationship between reading achievement and specific skills of spelling and reading.

This study compares *good* and *poor* readers at fourth- and eighth-grade levels on each of four measures. These are spelling, spelling of phonetic syllables, syllabication, and intelligence. The populations for the study consisted of the fourth and eighth grade pupils of the white schools of one of Georgia's 198 school districts. The investigation involved 175 pupils in the fourth grade and 155 pupils in the eighth grade. Throughout the study data for each grade level were treated separately. Testing was carried out during the seventh and eighth months of the school year.

Selecting the Good and Poor Readers

The California Reading Test, Form **DD**, was administered to all pupils in the fourth and eighth grades of the system involved. The Elementary test was used for the fourth-grade pupils, and the Intermediate level was administered to the eighth-grade pupils. By definition, *good* readers were those who scored in the top 27 per cent of their grade group on the reading test. *Poor* readers were those who scored in the bottom 27 per cent of their grade group.

The *poor* readers at the fourth-grade level included 47 children who scored at 4.2 grade placement level or below. The mean raw score for the *poor* readers was 64, which may be converted to a reading grade placement of 3.8. The 47 *good* readers, those in the top 27 per cent, scored at 5.6 grade placement or above. The mean raw score for this group was 107, and may be converted into a grade placement of 6.2.

The bottom 27 per cent of the eighth-grade group was made up of 42 pupils who had grade placement scores of 7.0 or below on reading. The top 27 per cent was made up of 42 pupils who had grade placement scores on reading of 9.0 or above. The mean raw score of the *poor* readers was 61, which may be interpreted from the test norms as a grade placement of 6.0. The mean raw score of the *good* readers was 116, which may be assigned a grade placement of 9.8.

Other Tests Used

A 60-word, list-dictation spelling test was administered to measure spelling achievement. It was based upon three tests previously developed and used by the writer (1). Twenty words were selected randomly from each of the three tests. The words were originally drawn from Ashbaugh's list (2) and were classified in accordance with phonetic content. Test I, from which 20 words were selected, consisted of words which were purely phonetic in composition, i.e., they were spelled as they sounded. Examples of these words are *thus*, *himself*, and *riding*. Test II, from which a second 20 words were taken, contained partially phonetic words, those deviating from the purely phonetic through such differences as having double letters with only one sounded, silent *e* at the end of words, and vowel digraphs with one letter silent. Examples of these words are *oats*, *pure*, and *load*. Test III, also contributing 20 words to the present test, was made up of words which were nonphonetic in at least one element. Examples of these are *says*, *washed*, and *island*. Words on the test ranged from the purely phonetic to the non-phonetic in at least one component. The same test was administered to both grade levels used in the study.

The spelling of phonetic syllables test was a test in which nonsense syllables were called out for pupils to spell just as they sounded. Examples are "tingure" and "chekzub." Any reasonable phonetic spelling was accepted as correct, and half credit was given for minor deviations from strictly phonetic spelling. Of a large battery of tests used in a previous study, this test had been found to be the best single predictor of spelling success (3).

The syllabication test contained nonsense words to be broken into syllables in accordance with four principles of syllabication which are usually taught as a part of reading instruction near the end of the primary grades. These involved principles of syllabication applying to the following four situations:

TABLE I

COMPARISONS OF 47 GOOD AND 47 POOR READERS ON FOUR MEASURES, FOURTH GRADE

Measure	Mean of Good	S. D. of Good	Mean of Poor	S. D. of Poor
Spelling	51.47	6.80	21.79	14.17
Spelling of Phonetic Syllables	30.53	12.04	10.34	10.06
Syllabication	13.51	2.23	10.62	3.47
Intelligence	88.02	9.58	72.47	12.56

TABLE II

COMPARISONS OF 42 GOOD AND 42 POOR READERS ON FOUR MEASURES, EIGHTH GRADE

Measure	Mean of Good	S. D. of Good	Mean of Poor	S. D. of Poor
Spelling	58.62	2.73	46.21	12.61
Spelling of Phonetic Syllables	43.31	7.56	26.60	11.61
Syllabication	15.62	3.15	12.74	2.39
Intelligence	108.00	9.29	87.48	14.10

TABLE III

GOOD READERS WHO WERE POOR ON OTHER MEASURES, FOURTH AND EIGHTH GRADES

Measure	47 Fourth Graders		42 Eighth Graders	
	No.	Per Cent*	No.	Per Cent*
Spelling	0	0	3	7
Spelling of Phonetic Syllables	2	4	1	2
Syllabication	4	9	5	12
Intelligence	5	11	4	10

*Per cent refers to proportion of pupils among 47 or 42 good readers and not to proportion of total number of pupils in study.

TABLE IV

POOR READERS WHO WERE GOOD ON OTHER MEASURES, FOURTH AND EIGHTH GRADES

Measure	47 Fourth Graders		42 Eighth Graders	
	No.	Per Cent*	No.	Per Cent*
Spelling	2	4	2	5
Spelling of Phonetic Syllables	3	6	2	5
Syllabication	13	28	5	12
Intelligence	5	11	0	0

*Per cent refers to proportion of pupils among 47 or 42 poor readers and not to proportion of total number of pupils in study.

a single consonant comes between two vowels; two consonants come between two vowels; a consonant is followed by le; and d or t is followed by ed. Words were passed out in mimeographed form, and pupils were asked to divide them into syllables. The list contained 25 words such as "tumil," "pulmir," "megle," and "winted."

The California Test of Mental Maturity, Non-Language section, was used to estimate intelligence.

The Hoyt Analysis of Variance method (4) was used for computing reliabilities on the three tests constructed by the writer. The reliabilities found were as follows: spelling, .97 for fourth grade and .96 for eighth grade; spelling of phonetic syllables, .97 for fourth grade and .96 for eighth grade; and syllabication, .54 for both fourth- and eighth-grade levels.

Comparing the Good and Poor Readers

A comparison of the 47 good and the 47 poor fourth-grade readers on each of the four measures was made. Table I shows that the two groups differed considerably on each of the four measures. The mean performance of the good readers was superior to the mean performance of the poor readers on all tests.

The 42 good and the 42 poor eighth-grade readers were also compared on each of the remaining four measures. Table II reveals that the two groups differed in mean performance on each of the four measures, the larger mean always being that of the good readers.

A second phase of the study involved examining the good and poor readers to determine whether or not they were uniformly good or poor on all measures. The fourth- and eighth-grade groups were divided into good and poor performers on each of the four other measures. Good and poor again meant being in the upper or lower 27 per cent. On each measure, these groups were made up of the upper 47 and the lower 47 pupils at the fourth-grade level and the upper 42 and the lower 42 pupils at the eighth-grade level.

Five fourth-grade pupils were in the top 27 per cent on all measures. Another five fourth-grade pupils were in the bottom 27 per cent on all tests. Twelve eighth-grade pupils were in the top 27 per cent on all measures, and eight eighth-grade pupils were in the bottom group on all tests.

Table III reveals that pupils in the top 27 per cent on reading achievement sometimes fell in the bottom 27 per cent on one or more of the other measures. This was found at both fourth- and eighth-grade levels. At least one pupil who performed well in reading scored among the lowest 27 per cent on each of the other measures, except in fourth-grade spelling achievement. Those fourth graders high in reading and low on one or more of the other measures ranged from no one in spelling to five pupils on intelligence. The eighth-grade deviations ranged from one case on spelling of phonetic syllables to five cases on syllabication.

Table IV shows the number of pupils in the bottom 27 per cent on reading achievement who were in the top 27 per cent on one or more of the other measures. Some poor readers were found to be good in performance on each of the other four measures at both grade levels except in one instance, that of intelligence in the eighth-grade group. These ranged from two pupils on spelling to thirteen pupils on syllabication in the fourth-grade group. The eighth-grade discrepancy cases ranged from no pupils in intelligence to five pupils on syllabication.

Summary and Conclusions

At both fourth- and eighth-grade levels, good readers differed significantly from poor readers on all other measures - spelling, spelling of phonetic syllables, syllabication, and intelligence.

Though good and poor readers differed considerably as groups on all of the other measures, some good readers did not perform well on one or more of the other measures. It was also found that some poor readers performed well on one or more of the other measures.

Insofar as these two groups are concerned, this investigation supports the contention that though certain skills or abilities may be associated with good reading performance, individual children may be good readers and still be deficient in some one or more of these skills. A child also may be strong in some particular reading technique or skill generally associated with success in reading and still be poor in reading. Conflicting opinions expressed by teachers may arise at least in some instances from the teachers' knowledge of children who deviate from the group finding.

FOOTNOTES

1. Aaron, I. E., "The Relationship of Auditory-Visual Discrimination to Spelling Ability." Unpublished Ph. D. thesis, University of Minnesota, 1954.
2. Ashbaugh, Ernest J., The Iowa Spelling Scales, College of Education, Journal of Education Research Monographs, No. 3, 1922.
3. Aaron, op. cit.
4. Hoyt, Cyril J., "Note on a Simplified Method of Computing Test Reliability," Educational and Psychological Measurement. 1:93-95, 1941.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 1, September 1960)

COMPARISON OF EDUCATION STUDENTS, SUCCESSFUL SCIENCE TEACHERS AND EDUCATIONAL ADMINISTRATORS ON THE EDWARDS PPS

REED M. MERRILL
University of Utah

THE PURPOSE of the research to be reported here is to compare several teacher groups on a personality test standardized on a college population and designed to measure the relative strength of 15 personality needs or dimensions. There are several reasons for exploring the need structure of occupational groups such as teachers: 1) Teachers are in a position to serve as models for individuals whose preferences are uncertain. With the exception of parents, teachers comprise the group of adults with whom the child has most frequent and intimate contact. 2) Teachers are relatively free to reward or punish, implicitly or explicitly, those preferences which children do exhibit. It is of concern to ask what sorts of preferences are likely to be approved or disapproved by persons exhibiting such a need structure. 3) Such data can at least be suggestive in the areas of teacher selection and training. This report is directed to the problem of describing and discussing the need structure of public school teachers as measured by the Edwards Personal Preference Schedule (PPS).

The PPS consists of 210 forced-choice items. Each pair of items is matched approximately for mean social desirability value, as previously judged by college students, to minimize the effect of social desirability on item choice. Items measuring one need are paired twice with each of the remaining 14 needs; thus the total raw score on the test will be the same for all persons. The needs measured by the PPS are those indicated in the manual (2) and follow Murray's (6) definitions and nomenclature.

The cases were grouped as follows: 1) 781 male students in the College of Education, University of Utah, who enrolled in an orientation course during 1955-56. This course is required of entering education students and most of the students were Freshmen or Sophomores. The average age was 20.4 years. In the report, this group is known as the "Education Student" group. 2) 53 Male teachers who were enrolled in an Academic-Year Institute in the science-teaching field supported by the National Science Foundation, 1957-58. This group has an average teaching experience of seven years and were judged to be successful by their supervisors. The average age was 34 years. This group is known as the "Successful Teacher" group. 3) 32 School administrators who were involved in an in-service training program. This group had an average of ten years experience in the schools and were presently employed as administrators. The average age was 38.4 years. This group is designated "Administrators."

Table I presents the comparison of the groups with the norm group of college males on the PPS. As can be seen in this table, if these groups are looked upon as randomly selected groups, which strictly speaking they are not, there are a number of statistically significant differences. The education student group is more deferring, orderly, affiliative, intraceptive, abasing, nurturant, enduring, and less autonomous, succorant, change-oriented, heterosexual and aggressive than the norm group in need structure as measured by the PPS. The successful teacher group is more deferring, orderly, abasing, enduring and less exhibitionistic and heterosexual than the norm group. The administrator group is more deferring, orderly, enduring and less exhibitionistic, autonomous, and heterosexual than the norm group.

Since one-half of the education student group do not stay in teaching, perhaps the differences obtained with this group relate to needs that tend to take education students away from teaching, e.g., the student group is significantly more exhibitionistic than mature teachers or administrators. Often teaching is couched in terms of aiding and assisting others, interaction in many groups, and analyzing the behavior of others. If this is so, one would expect high nurturance, affiliation and intraception scores. Such is not the case. The present findings are consistent with previous work by Barr (1), Jackson and Guba (5) and Fulkerson (4).

The evidence seems to indicate that teachers are not highly motivated by a strong interest in social service, by powerful nurturant needs, or even by a deep interest in children. Descriptively, the suc-

TABLE I

COMPARISON OF EDUCATION STUDENTS, SUCCESSFUL TEACHERS, AND EDUCATIONAL ADMINISTRATORS WITH THE NORM GROUP OF COLLEGE MALES ON THE EDWARDS PPS

Need	Norm Group (N = 760)		Education Students (N = 781)		Mature Teachers (N = 53)		Educational Administrators (N = 32)	
	Mean	S.D.	Mean	S.D.	Mean	S. D.	Mean	S.D.
Achievement	15.66	4.13	15.37	4.09	16.77	4.04	15.63	3.41
Deference	11.21	3.59	13.56**	3.69	14.17***	4.21	15.41**	3.44
Order	10.23	4.31	11.72***	4.15	12.41***	4.23	13.03***	3.19
Exhibition	14.40	3.53	14.64	3.64	13.23**	3.75	12.44*	4.38
Autonomy	14.34	4.45	12.40***	4.28	13.24	4.32	11.91***	3.50
Affiliation	15.00	4.32	15.66**	4.30	15.62	4.23	16.09	3.34
Intraception	16.12	5.23	17.37***	4.79	15.00	5.09	16.91	5.84
Succorance	10.64	4.70	9.65***	4.52	9.92	3.83	10.28	4.44
Dominance	17.33	4.88	17.65	4.45	16.68	4.68	17.41	4.22
Abasement	12.24	4.93	13.82***	4.77	13.94**	4.34	12.38	3.84
Nurturance	14.04	4.80	14.73**	4.79	13.83	4.43	15.44	5.03
Change	15.51	4.74	15.04*	4.35	16.04	4.04	13.91	5.13
Endurance	12.66	5.30	14.08***	4.98	16.24***	5.57	15.16***	3.50
Heterosexuality	17.65	5.48	13.35***	5.85	10.89***	5.66	12.25***	5.38
Agression	12.79	4.59	10.77***	4.51	12.00	4.11	11.66	4.72

* Difference with norms significant at the 5% level of confidence.

** Difference with norms significant at the 1% level of confidence.

*** Difference with norms significant at the .001% level of confidence.

cessful teacher groups appear obsequious, eternally patient, painstakingly demanding and with less achievement and individual drive than other professional groups. These results are strikingly similar to Jackson's and Guba's (5) findings with 366 public school teachers in the Chicago area. These findings suggest that such a "meek" need structure for teachers may be related to the selection process for this occupational group. Conceivably this could also be a factor in the teacher-pupil pattern of learning that has developed in our schools. Further research is required on this point since the findigs in this study are essentially descriptive.

In summary, this paper presents an attempt to describe the manifest need structure of teacher groups as measured by the Edwards PPS. Needs found to be most characteristic of successful teachers or administrators were high deference, order and endurance and low heterosexuality and exhibition. Marked variation from this pattern was found for education students and the differences may relate to the fact that many education students do not stay in teaching. The pattern for successful teachers emerges with a strong "marketing orientation" (3). These findings are sufficiently encouraging to warrant further research with the PPS in relation to the general problems of teacher selection and training.

REFERENCES

1. Barr, Arvil S., Eustic, David E. and Noe, Edward J. "The Measurement and Prediction of Teacher Efficiency," *Review of Educational Research*, 25:261-296, 1955.
2. Edwards, Allen. *Edwards Personal Preference Schedule*. New York: The Psychological Corporation, 1954.
3. Fromm, E. *The Art of Loving*, New York: Harper Brothers, 1956
4. Fulkerson, Glen. "A Resume of Current Teacher Personnel Research," *Journal of Educational Research*, 47:669-682, 1954.
5. Jackson, Phillip W. and Guba, Egon G. *The Need Structure of In-Service Teachers: An Occupational Analysis*, University of Chicago, 1957, 1-28. (Mimeographed Report).
6. Murray, Henry A. *Explorations in Personality*. New York: Oxford University Press, 1938.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 OCTOBER, 1960 NUMBER 2

CONTENTS

Characteristics of Slow Learners and Pupils of Normal Intelligence Referred to a Child Guidance Clinic—Walter I. Murray and Irving Bloom 43

College Students and Their Assignments—Janet Bassett Baker 49

Men Teachers in Public Elementary Schools—Robert S. Thurman 54

Use of Results of Large-Scale Testing Programs in Instruction and Guidance—Arthur E. Traxler 59

The Assertion of Authority: Differences Between Teachers, Student-Teachers, and Mothers of Young Children—Albert S. Dreyer and Dorothy Haupt 63

A Survey of Educational Disability in Emotionally Disturbed Children—Arthur S. Tamkin 67

Experiment in Cross Grouping—Jeff West and Callie Sievers 70

Counselor and Student Estimates of Grades as Predictors of Academic Achievement—Malcolm H. Robertson 73

The Relationship of Certain Factors to Marking Practices in Individualized Reporting Programs—Joseph W. Halliwell 76

Social Status and Enrollment in Traditional Mathematics—Marlow A. Markert 79

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

CHARACTERISTICS OF SLOW LEARNERS AND PUPILS OF NORMAL INTELLIGENCE REFERRED TO A CHILD GUIDANCE CLINIC

WALTER I. MURRAY, Brooklyn College
IRVING BLOOM, New Jersey State Teachers College, Newark

MANY OF THE CHILDREN referred to the Bureau of Child Guidance of the New York City Public Schools are considered slow learners because they score below the normal range in I.Q., that is, between 75 and 90 on individual intelligence tests. Often others who score within the normal range of 91 and 110 are considered slow learners by school personnel and parents. The purpose of the study is to determine what similarities and differences exist between these two groups.

Definition and Selection of Cases

The data regarding the characteristics investigated are contained in the Face Closing Sheets filled out for all cases accepted for study and/or treatment by the Bureau of Child Guidance and closed when Bureau activity is completed, regardless of the reason for completion. The characteristics of each case have been classified by Bureau of Child Guidance staff members according to the Manual of Recording Procedures for Statistical Reports, 1954. This material is codified and the data are punched on I.B.M. cards. The sampling consists of all cases which fall between the I.Q. ranges of 75 to 90 to 110, which were selected from the total of all cases which were closed in the Bronx Center of the Bureau during the school year 1954-55.

Method of Procedure

The I.B.M. cards were sorted into two groups; group A consisted of all cases which fall within the 91 to 110 I.Q. range. The number and percentage of cases in each group, in terms of characteristics described as follows, were computed and comparisons were made between them:

1. Problems as referred
2. Age at referral
3. Number of siblings
4. Family status
5. Children showing major physical defects
6. Number of problems by each child
7. Reading grade
8. Arithmetic grade

Differences between groups on each of the factors studied were calculated and tests for significance of differences were made through the use of the Chi-square statistic. This statistic is used to test the hypothesis that the samples have been drawn from a population with a parameter equal to a theoretical ratio. It also tests the hypothesis that any difference between the groups in terms of observed frequencies can be accounted for in terms of sampling variation. If the obtained value of Chi-square is such that it will occur 5 percent of the time or less, then, according to arbitrary standards, it is assumed that the hypothesis being tested is untenable.[1]

Value of the Study

It is hoped that the findings will give teachers a better understanding of slow learners and, in particular, the educationally retarded pupil of normal intelligence. It is also hoped that such understanding may have a practical application to practices and procedures in teaching.

Findings

Problem as referred. Table I indicates the number of subjects of each group who were referred to the Bureau of Child Guidance on the basis of specific problems as observed by personnel of the school in which children were enrolled. The problems are behaviorally defined as follows:

1. Resistance to going to school--includes general objection to going to school, school phobias, truancy, etc.
2. Educational Guidance--includes classification, vocational guidance.
3. Suspected retardation.

---

* Footnotes will be found at the end of the article.

TABLE I

FREQUENCY OF EACH PROBLEM

Problem	Number of Children		Percent of Total Children			
	High IQ	Low IQ	High IQ	Low IQ	High IQ	Low IQ
Resistance to Going to School	18	8	11.6	2.1	15,1	2.4
Educational Guidance	34	95	21.9	25.1	28.6	28.0
Suspected Retardation	37	208	23.9	54.9	31.1	61.4
Behavior Difficulties	26	24	16.8	6.3	21.8	7.1
Anti-Social Behavior	5	8	3.2	2.1	4.2	2.4
Withdrawn Behavior	14	22	9.0	5.8	11.8	6.5
Symptomatic Habit Formation	5	7	3.2	1.9	4.2	2.1
Anxiety Disturbance	10	5	6.5	1.3	8.4	1.5
Expiration of Exemption	2	2	1.3	.5	1.7	.6
Epilepsy	4	-	2.6	-	3.4	-
Total	155	379	100.0	100.0	---	---

4. Behavior difficulties--includes disruptive class procedure, overly árgumentative and negativistic behavior in home and school, restlessness and overactivity, temper tantrums, attention-getting behavior, etc.

5. Anti-social behavior--includes lying, cheating, stealing, destroying property, turning in false alarms, running away from home, etc.

6. Withdrawn behavior--includes shyness, timidity, day dreaming, uncommunicativeness, unhappiness, etc.

7. Symptomatic habit formation--includes chewing fingers or objects, thumb sucking, nail biting, enuresis, tic-like movements, soiling, masturbation, etc.

8. Anxiety disturbance--includes fears, tension, worries, etc.

9. Expiration of exemption--includes re-examination for exemption.

10. Epilepsy.

In comparing the two groups with respect to problem as referred, we find that there is a statistically significant difference between the two groups in that the high I.Q. group has more problems of this nature (Chi-square equals 7.49; p is less than .01).[2]

It is apparent that certain problems have greater incidence than others. For example, 37 (23.9%) of the high I.Q. groups were referred because of suspected retardation, while 208 or 54.9% of the low I.Q. group were referred for this reason. This large difference between the two groups supports the belief that children whose I.Q.'s are relatively low are likely to be more academically retarded than children who are within the normal I.Q. range.

Age at referral. There is no statistically significant difference between the groups with respect to age at date of referral. The greatest numbers of children in both groups fall between the ages of seven and thirteen, normally the age range of the elementary school.

Number of siblings. The total number of siblings and the average number of siblings in both groups were compared, with no statistically significant difference between the two groups found. Data on this characteristic of the group were incomplete as reflected in the large number of children in both groups for whom this information was not given.

Social findings. Table II indicates the social findings of the two groups. The groups were compared with respect to aspects of their social life. There are statistically significant differences between the two groups (Chi-square equals 40.9; p is less than .001). This is interpreted to indicate that when the I.Q. is high, the group has more social problems than the low I.Q. group. However, we had multi-scoring; this does not mean that more children of the high I.Q. group had problems, but that the total number of social problems for the high I.Q. group was greater. The greatest number of problems for both groups clustered around problems which had to do with family relationships. The problem of greatest frequency in this area had to do with disturbed parent-child relations.

Family status. Data on family status were sparse. There were 35 children (29.4%) of the high I.Q. group, and 59 (17.4%) of the low I.Q. group living with both parents. It should be pointed out that more than one-half of the high I.Q. group, and approximately three-fourths of the low I.Q. group, had no data on family status.

Major physical defects. Relatively few children in either group have records of major physical defects; only 12 of the high I.Q. group and 9 children of the low I.Q. group. It is difficult to hypothesize about this circumstance.

Number of problems. In an effort to obtain deeper insights about the two groups, an analysis of the problems possessed by each child was made. Table III presents these data. More than three-fourths of the children of each group were identified as having only one problem. Even though there are significant differences between the groups with respect to the number of problems (Chi-square equals 38.4; p is less than .001), one cannot evaluate these differences qualitatively. The problems were in more than one direction and of varying degrees of seriousness. More important, a child with one problem may be in more serious difficulty than another child with three problems of a less serious nature. It can be pointed out that the number of children with three problems was relatively small in each group.

Reading grade. Table IV presents the number of children in both groups who made scores on reading tests indicating grade placements of the children. The high I.Q. group has better reading scores than the low I.Q. group. The difference in performance on this test was statistically significant (Chi-square equals 14.3; and p is less than .01). It should be pointed out that about 50% of each group had reading scores which indicated placement in primary grades. In considering the chronological age of these children in terms of their reading performance, we find that approximately one-half of the children are in the primary grades. The data presented are of such a nature that it is impossible to delineate the exact nature of the grade level because of the grossness of the tabulating procedures.

Arithmetic grade. Table V presents the number of children who made scores on arithmetic tests and the grade placement indicated by the scores. As in the case of reading performance, there was statistically significant difference in favor of the high I.Q. group (Chi-square equals 11.6; p is less than .01). The number of children whose scores indicated achievement at a primary grade level was about the same as that of reading. Teachers often refer children for guidance and counseling when there is strong suspicion and supporting evidence of reading and arithmetic disabilities.

TABLE II

FREQUENCY OF SOCIAL FINDINGS

	Number of Children		Percent of Total Children			
	High IQ	Low IQ	High IQ	Low IQ	High IQ	Low IQ
Family Economic Problem Affecting Child						
Under-Employment	5	6	4.3	4.1	4.2	1.8
Unemployment	4	8	3.4	5.5	3.4	2.4
Budgetary Defects	7	8	6.0	5.5	5.9	2.4
Family Housing Problems						
Standard	4	6	3.4	4.1	3.4	1.8
Overcrowding	8	16	6.8	11.1	6.7	4.7
Family Relationship						
Disturbed Parent-Child Relation	37	47	31.6	32.4	31.1	13.9
Broken Home	7	14	6.0	9.7	5.9	4.1
Unmarried Parenthood	7	4	6.0	2.8	5.9	1.2
Sibling Rivalry	11	9	9.4	6.2	9.2	2.7
Others	1	-	.9	---	.8	---
Illness Affecting Child						
Physical	8	5	6.8	3.4	6.7	1.5
Mental	1	4	.9	2.8	.8	1.2
Indication for Child Placement						
Emotional	10	8	8.5	5.5	8.4	2.4
Practical Reason	-	2	---	1.4	---	.6
Neighborhood Facilities						
Lack of Recreational Facilities	-	2	---	1.4	---	.6
Delinquency	7	6	6.0	4.1	5.9	1.8
Total	117	145	100.0	100.0		
No Findings	69	272			58.0	42.8

TABLE III

NUMBER OF PROBLEMS BY EACH CHILD

Number of Problems	Number of Children		Percent of Total	
	High IQ	Low IQ	High IQ	Low IQ
1 Problem	91	302	76.5	89.1
2 Problems	20	34	16.8	10.0
3 Problems	8	3	6.7	.9
Total Children	119	339	100.0	100.0

TABLE IV

READING GRADE

Grade	Number of Children		Percent of Total	
	High IQ	Low IQ	High IQ	Low IQ
Pre-School	9	62	7.6	18.3
1 - 3	59	168	49.6	49.5
4 - 6	12	29	10.1	8.6
7 and Over	6	2	5.0	.6
None of the Above	33	78	27.7	23.0
Total	119	339	100.0	100.0

TABLE V

ARITHMETIC GRADE

Grade	Number of Children		Percent of Total	
	High IQ	Low IQ	High IQ	Low IQ
Pre-School	4	31	3.4	9.2
1 - 3	55	179	46.2	52.8
4 - 6	22	32	18.5	9.4
7 and Over	1	3	.8	.9
None of the Above	37	94	31.1	27.7
Total	119	339	100.0	100.0

Conclusions

In summary, the purpose of this paper is to determine what similarities and differences exist between two groups of children who were referred to the Bureau of Child Guidance because they were slow learners. Their I.Q. ranges were 75 - 90 and 91-110 respectively. On the basis of the data afforded, one may observe that there were statistically significant differences between the high I.Q. group and the low I.Q. group with respect to some characteristics: problems as referred, social findings, number of problems, reading grade and arithmetic grade.

The differences were in favor of the high I.Q. group. In some characteristics there were no statistically significant differences found between the groups: age at referral, number of siblings, family status and major physical defects.

Incomplete individual records made it impossible to derive more meaningful and conclusive inferences. This shows that the mere presence of blanks on data forms will not insure their use.

The evidence in this study indicated that children are referred to the Bureau of Child Guidance for several reasons. Suspected retardation and social factors appeared most frequently as bases for referral.

FOOTNOTES

1. Edwards, Allen L., Statistical Analysis, New York; Rhinehart and Company, Inc., pp. 239-261, 1956.
2. This indicates that the difference in problems between the two groups would occur on the basis of chance less than one out of one hundred times or at the 1% level of significance. Whenever statistical differences are found through the use of this statistic in the remaining portion of this report, it should be interpreted similarly.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

COLLEGE STUDENTS AND THEIR ASSIGNMENTS

JANET BASSETT BAKER
Marywood College, Scranton, Pennsylvania

COLLEGE ASSIGNMENTS can be made to stimulate students' interests, to promote learning, and among many things to provide material for research. Consideration is given in this article to: 1) conclusions already reached in the assignment controversy; 2) college students' experiences with assignments; 3) college students' comments on assignments; 4) suggestions for individualized assignments on the collegiate level; and 5) conclusions.

First, what conclusions have already been reached in the outside-of-the-class work controversy? According to Benjamin Fine (7) in the New York Times Magazine, the problem of assignments originated in 1904 when a German educator came to the conclusion that homework hindered rather than helped the school work of twelve- and thirteen-year-old pupils. Fine mentioned other studies which had been made in Tennessee and Vermont. Both of these studies agreed with the findings of the German educator. What is more, Fine maintained that students who attended colleges where no outside assignments were given did better than those who went to colleges where the traditional assignments were made.

Another research study, made in an Oklahoma high school, showed that pupils who were assigned work which supplemented the regular class work had a better sense of direction than those who had "no" homework. Of significance is the conclusion: "Home study properly assigned and evaluated so far as it relates to the pupils in this experiment is an aid in improving scholarship."(2)

According to a certain author, there is only one thing upon which proponents and opponents in the homework controversy are in accord, namely, "the growing young man or woman should be active during his after-school hours in worthwhile activities." (4) The boy practices all afternoon to make the varsity; the young man spends the entire day in the library to prepare for the debate; and members in the school orchestras and bands rehearse for hours. Since these activities are all work outside the regular classroom, does not the real problem pertinent to homework become a matter of the teachers' presenting lessons or delivering lectures so rich and varied that the students will want to do additional work on their own? (5)

Second, what are college students' experiences with assignments; that is, which experiences help and which hinder their scholastic achievement?

To secure the necessary data for answering this question, a brief questionnaire was given, early in September, 1956, to 102 women at the College of Notre Dame, Baltimore, Maryland; and to 6 women and 70 men at the Loyola Evening College and Graduate Division, Baltimore, Maryland. Later, early in September, 1958, it was administered in the Baltimore Junior College to 14 women and 82 men and to 77 women at Marywood College, Scranton, Pennsylvania. Of the total 351 college students, 199 were women and 152 were men.

In Part I of the questionnaire, the students were asked to check the one experience which helped them most to keep up their grades and to develop their capacities and abilities in any of their high school or college courses. There was, also, provision for comments on any helpful aspects of their assignments.

In Part II of the questionnaire, the students were asked to check all experiences with their assignments in any subject on the secondary or collegiate level which had hindered their progress. Comments were also welcomed. The questionnaire is as follows:

Part I: College Students' Helpful Experiences With Assignments

Directions: In each group, check the one most helpful experience with your assignment. Write comment in space provided.

GROUP I. WHEN the assignment was given.
____A. At the beginning of the semester.
____B. Before, during, or after each lecture.
____C. Before each major division of the subject.

GROUP II. When the TIME required for doing the assignment was
____A. Left entirely to the student.
____B. Put at a minimum of two hours on the assignment for each hour spent in the class.
____C. Put a maximum of two hours on each nightly assignment.

GROUP III. When the assignment CONSISTED OF
____A. Certain chapters in one or more good texts.
____B. Collateral reading in primary sources.
____C. Outside-of-the-book assignments: trips, parts in plays, TV, or panel; sitting in or interning in legislative and judicial bodies; independent selection by students.

Part II: College Students' Experiences of Hindering or Retardation Caused by Assignments

Directions: Check all experiences which have hindered or delayed your mastering a subject. Write comments in space provided.

____A. Whole class was assigned one reference when there were only two or three copies available.
____B. Instructor's objectives were indefinite.
____C. Change was made in the assignment whenever the instructor was requested to do so by a few of his students.
____D. Preparation was needed for two or three informal tests scheduled for the same day in different courses.
____E. Term paper or research paper was due, or informal test was announced for Monday or on the day following a vacation, or on the day following a special activity in college.
____F. Written assignments were not returned to students.

The basic results for the college student's experience with assignments are presented in Tables I through IV. An examination of Tables I and II shows that a majority of all the students, as well as a majority of those in each college, experienced most stimulation and help in mastering a subject when the assignments were made at the beginning of the semester; yet almost a third of those who expressed their experiences showed that they derived most benefit from an assignment which was made before, during, or after each lecture; and over a tenth benefitted most from an assignment made before each major division of a subject. Variations exist between the different colleges as to the advisability of giving assignments before, during, or after each lecture, and also before each major division.

Tables I and II show that a majority of all the participants in this study, as well as a majority in each separate college, want to decide for themselves how much time they should spend on an assignment. Wide variations present themselves in Tables I and II with respect to the students' putting a minimum or a maximum number of hours, which are set by the colleges, on the outside-of-class work.

Table I shows that over three-fifths of all the students secured most help from an assignment consisting of certain chapters in one or more good texts. Yet over a quarter of them benefited most from collateral reading in primary sources, and over an eighth were benefited most by outside-the-book assignments. Wide variations exist among the separate colleges as to the assistance which students derived from assignments in the primary sources or in assignments that embraced outside-the-book work.

The statistical data on Tables III and IV show that all the students who participated in this research project had experienced annoyance with assignments in high school or college when the whole class was referred to one reference and there were only two or three copies available; when the instructor's objectives or standards were indefinite; when change was made in the assignment whenever the instructor was requested by a few of his students' when the students had to study for several tests which were given on the same day; when papers were required at inopportune times; and when written assignments were not marked and returned (3).

Third, what comments did college students make about experiences with their assignments? Only two of the 351 students said "I do not receive any help from work done outside the class." There were 81 students who made remarks similar to the following: "I learn much from outside work which is a continuation of the current lecture; I get a real grasp of the subject when I go over my notes; I impress the subject upon my mind when I read about it in the primary sources which I have come to enjoy reading; subject means more to me as I get deeper and deeper into it by extra study; and I'm paying my hard-earned money for these courses, and I want plenty of outside study."

One woman who had returned to college after an absence of ten years said "Don't take the outside work away from me. I would be lost without it." A man with a large family to support asserted "I enjoy reading and studying outside class. These are my chief recreations." Even a high school teacher who was attending college at night confided "Stop all this chatter about assignments and the help afforded by outside study. Pile it on, and be sure to see that we do it."

Consequently, it appears that college students want outside work, and it is up to the college instructor to stimulate more students to acquire the 'itch to know.'(6) There is only confirmation in this study of Ahrens' student's remark "I like homework . . . . I get rattled and make mistakes when I have to work fast, but at home I can take my time and do the job right." (1)

Fourth, what provision did the conductor of this study make for assignments to improve students' independent study? She prepared a syllabus for each of her three separate preparations. Each syllabus gave the objectives which the students expected to have attained at the completion of the course. For example, one of the goals for the Modern European History, 1500 to the Present, is to have students become interested in historical reading and to have them appraise good historical writing critically.

TABLE I

MOST HELPFUL ASSIGNMENT EXPERIENCES FOR 351 COLLEGE STUDENTS

Group	Most Helpful Experience	Number of Cases	Percentages
I	A	189	53.8
I	B	114	32.5
I	C	48	13.7
II	A	275	78.3
II	B	50	14.2
II	C	26	7.4
III	A	217	61.8
III	B	90	25.6
III	C	44	12.5

TABLE II

MOST HELPFUL ASSIGNMENT EXPERIENCES FOR STUDENTS IN EACH COLLEGE

Group	Most Helpful Experience	102 Students Notre Dame, Baltimore		76 Students Loyola, Baltimore		96 Students, Junior College, Baltimore		77 Students Marywood, Scranton	
		Cases	%	Cases	%	Cases	%	Cases	%
I	A	60	58.8	62	8.16	45	46.9	38	49.4
I	B	36	35.3	8	10.5	29	30.2	25	32.4
I	C	6	5.9	6	7.9	22	22.9	14	18.2
II	A	81	79.4	61	80.3	74	77.1	59	76.6
II	B	20	19.6	14	18.4	8	8.3	8	10.4
II	C	1	.99	1	1.3	14	14.6	10	13.0
III	A	60	58.8	40	81.6	60	62.5	57	74.0
III	B	36	35.3	26	34.2	14	14.6	14	18.2
III	C	6	5.9	10	13.2	22	22.9	6	7.8

TABLE III

ASSIGNMENT EXPERIENCES WHICH HINDERED 351 COLLEGE STUDENTS

The Hindering Experiences	Cases	Percentages
A	285	81.2
B	254	72.4
C	144	41.0
D	282	80.3
E	249	70.9
F	268	76.4

TABLE IV

ASSIGNMENT EXPERIENCES WHICH HINDERED STUDENTS IN EACH COLLEGE

Hindering Experience	102 Students Notre Dame, Baltimore		76 Students Loyola, Baltimore		96 Students Junior College, Baltimore		77 Students Marywood, Scranton	
	Cases	%	Cases	%	Cases	%	Cases	%
A	88	86.3	70	92.1	62	64.6	65	84.4
B	81	79.4	62	81.6	47	49.0	64	83.1
C	40	39.2	61	80.3	20	20.8	23	29.9
D	89	87.2	72	94.7	65	67.7	56	72.7
E	81	79.4	68	89.5	54	56.3	46	59.7
F	88	86.3	70	92.1	52	54.2	58	75.3

The Modern European History syllabus gave the assignment for every week in the year from September through May. For instance, during the week of September 29, the topic was "Religious Revolt and Strong Monarchy in Britain." Other than chapter assignments in several good college texts, any one of which the students were at liberty to choose, it was suggested that they consult the primary sources in such books as Raymond Phineas Stearns, Pageant of Europe, New York, Harcourt, Brace and Company, 1947, chapter 10. Several of the students had purchased this book, and there were copies in the college library. For students who wished to do another kind of independent outside study, the syllabus included a lengthy annotated bibliography. Among the books on this background reading were Charles Andrew Brady, Stage of Fools, New York, Dutton, 1953, p. 381; Garrett Mattingly, Catherine of Aragon, Boston, Brown and Company, 1944, p. 435.

Other than giving the weekly assignments, the Modern European History syllabus listed the dates for the informal tests and for the final examinations.

At the beginning of each week's assignment, the instructor inquired, "What's the topic for today's lesson? What preparation have you made for it?" If a student had done pertinent reading in sources or background books, she was encouraged to report on it briefly to her class. Students were asked to come to the instructor's office to discuss their outside-of-class work or to talk over their written work.

An entirely different kind of syllabus was prepared for the Junior Reading List Course on "The American Foreign Policy." Other than several good texts such as Samuel Flagg Bemis, A Diplomatic History of the United States, New York, Henry Holt and Company, 1957, the Juniors delved into the primary sources on the American Foreign Policy; they interviewed specialists in the field; they went to newspaper offices where they read old papers; they made snapshots of places in local history which had any bearing on the oral or written reports they were preparing; and they attended lectures which related to the country's foreign relations.

The syllabus for the senior seminar listed a dozen books on historiography. Once the seniors had gained an acquaintance with how to write history, they were given a choice of a "local" history topic for their thesis. A few of these were as follows: Attitude of the People of ________ to World War I or to World War II; History of an Industry in or near ________; or history of a local celebrity.

Once the seniors had chosen a topic for their senior seminar theses, they visited historical societies and examined their archives, wrote to persons who had original documents, interviewed everybody who had first-hand information on their topic, made pictures, and came continually to the instructor's office to report on what material they had found or to ask how to use the abundance of diaries, letters, and sources they had received.

Fifth, with respect to college students and their assignments, what conclusions are reached?

1. College students want out-of-class work because their experiences teach them that it aids them to achieve a higher standard of work in college.

2. Students, for the most part, prefer their assignments to be given at the beginning of the semester; however, the instructor will have to be persistently checking on the assignment. Reports to class, tests, examinations, and interviews with students are a few of the checking devices.

3. Most of the students want the time that is to be devoted to an assignment to be left to them, for decisions, in which case the instructor will be obligated to check on the assignment. A friendly chat on the campus or a round-table discussion are suggestions for informal checks. It is doubtful whether a quarterly and the final examination are adequate for checks.

4. College students prefer chapter assignments in many good texts; however, primary sources are also desired by the students. A variety of out-of-the-book assignments are necessary to individualize the student's home study which he does independently.

5. All annoyances which college students have hitherto experienced should be eradicated.

6. A syllabus, plus a variety of independent activities by students, will provide impetus to college students' assuming independent study. Such a suggestive guide, along with many informal conferences with the instructor, should help the students' learning.

7. Assignments are not to be regarded as meaningless chores, dreaded by both student and instructor. College assignments should so awaken students' interest that they promote scholarship, and if assignments do not, then something should be found to replace them.

REFERENCES

1. Ahrens, Carsten, "What Students Think About Homework," National Education Association Journal, 46: p. 373, September, 1957.
2. Anderson, William Ewart, "An Attempt Through the Use of Experimental Techniques to Determine the Effect of Home Assignments Upon Scholastic Success," Journal of Educational Research, 40: pp. 141-3, October, 1946.
3. Baker, Janet Bassett, "Using Examinations to Promote Learning," Improving College and University Teaching, Vol. VI, Autumn, 1958. pp. 122-5.
4. Boodish, Hyman M., "The Teacher's Page," Social Studies, 43, p. 170, April, 1952.
5. Ibid., p. 171.
6. Cain, Victor R., "Another Slant on the Value of Homework," The Clearing House, Vol. 30, p. 162, November, 1955.
7. Fine, Benjamin, "No More Homework? The Pros and Cons," New York Times Magazine, January 13, 1952, pp. 16, 39, 41.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

MEN TEACHERS IN PUBLIC ELEMENTARY SCHOOLS*

ROBERT S. THURMAN
Kansas State Teachers College
Emporia, Kansas

THIS STUDY was an examination of the role and status of men teachers in public elementary schools in the United States and of the problems peculiar to these men. It included (1) a study of personal and professional data obtained from a representative sampling of men elementary school teachers; (2) an examination of the opinions of a representative sampling of superintendents, men elementary school teachers, presidents of Parent-Teacher Associations, and a group of other specially selected educators concerning the role and status of men teachers in elementary schools; (3) a determination of the unique contributions, if any, that men were thought to make as elementary teachers; and (4) a determination of the problems, if any, that might discourage men from entering or remaining in elementary education. This study was confined to men teaching in grades one through six in public schools in the United States.

The study was made following a twelve-year period of exhortation on the part of numerous authors who urged men to take a more active role in the rearing of young children and in the education of children of elementary school age. It was during this period--1945 to 1957--that the number of men elementary school teachers more than doubled. This increase was particularly impressive in light of a prediction made in 1947 that men teachers would soon disappear from the elementary schools.

Methods of Procedure

To gather the data for the study, two instruments--an opinionnaire and a questionnaire--were used. The opinionnaire was devised to determine views of various groups regarding the role and status of men teachers in elementary schools. It was submitted to superintendents, men elementary school teachers, and presidents of Parent-Teacher Associations in 346 public school systems, representing a 5 per cent stratified random sampling of public school systems of the United States. The opinionnaire was also submitted to 138 other educators who were selected from elementary school principals, presidents of local branches of the Association for Childhood Education International, and college teachers who had either served as officers or as committee members in the Association for Childhood Education International. Table I shows the distribution of opinionnaires sent and returned. The data gained from the opinionnaire were treated in three ways. First, the analysis of variance was used to test whether or not the responses made by participants in the stratified random sampling of superintendents, men elementary school teachers, and PTA presidents differed significantly. Second, the responses made by the superintendents, men elementary school teachers, PTA presidents, and other selected educators were compared on a percentage basis. Third, replies by the participants to open-end statements included in the opinionnaire were summarized and analyzed.

In addition to the opinionnaire, a second instrument, a questionnaire, was sent only to men elementary school teachers participating in this study and was used to gather personal and professional information.

Findings of the Study

The men elementary school teachers participating in this study were, on the average, 34 years of age, married, had at least one child, and had served in the military forces.

In terms of professional preparation, this group of representative men teachers were better qualified than the average elementary school teacher. Nearly all of them were college graduates and almost three-fourths were studying at the graduate level or had received graduate degrees. Almost all of them had taken course work in elementary education and a large majority had had student teaching at the elementary school level, which was predominantly in the intermediate grades. A little more than one-tenth of the men had had no student teaching experience at any level.

---

*Unpublished Ed. D. dissertation, George Peabody College for Teachers, 1959.

TABLE I

DISTRIBUTION OF OPINIONNAIRES SENT TO AND RETURNED BY GROUPS SAMPLED AND BY OTHER SELECTED EDUCATORS

Opinionnaires Sent to	No. in School Systems		Number Sent Out	Usable Returns	
	In U.S.A.	In Sample		Number	Per Cent
Groups Sampled					
Northeast	1630	82			
Supt.			82	52	63
Men Tchrs.			649	310	48
PTA Pres.			70	23	33
Subtotal			801	385	48
Southeast	1423	71			
Supt.			71	54	76
Men Tchrs.			95	41	43
PTA Pres.			38	16	42
Subtotal			204	111	54
North Central	1860	93			
Supt.			93	77	83
Men Tchrs.			333	201	60
PTA Pres.			55	18	33
Subtotal			481	296	61
Plains	1136	56			
Supt.			56	40	71
Men Tchrs.			70	35	50
PTA Pres.			30	12	40
Subtotal			156	87	56
Western	872	44			
Supt.			44	36	82
Men Tchrs.			232	132	57
PTA Pres.			29	12	41
Subtotal			305	180	59
Other Selected Educators			138	115	83
Total	6921	346	2085	1174	56
Groups Sampled		346	1947	1059	54
Supt.			346	259	75
Men Tchrs.			1379	719	52
PTA Pres.			222	81	36
Other Selected Educators			138	115	83

Most of the men teachers participating in this study were elementary classroom teachers, but one-eighth of them were teaching principals. The great majority of these men taught children in the intermediate grades although many of them also taught primary-grade children in such special subject areas as art, music, or physical education. Only about 1 per cent of the men were full-time classroom teachers in the primary grades. It was surprising to find that such a small percentage of men were primary-grade teachers since approximately one-half of the superintendents, men teachers, PTA presidents, and other educators recommended that men be assigned to all grade levels in the elementary school. Furthermore, approximately one-fourth of all the participants believed that boys and girls needed a man teacher for at least one year during the primary grades. One possible reason for so few men teaching below the fourth grade could be that, as a majority of the participants in each group agreed, men had neither the preparation nor the desire to teach the younger children. Whatever the reason for so few men being assigned to the primary grades, the contributions attributed to men teachers were being made almost entirely at the intermediate level.

It was interesting to find that most of the men teachers were satisfied that no advantage had been taken of them in assignments of building, classroom, children, grade level, or jobs for which they received no extra compensation and where other teachers were not assigned comparable duties. Furthermore, a majority of the superintendents, men elementary school teachers, PTA presidents, and other educators believed that men should not be given assignments differing from those given to women merely on the basis of sex, but that assignments should be made on the basis of individual ability.

The average salary received by this national sampling of men elementary school teachers was $4370, although the average received by men teaching in the Southeast and the Plains was approximately $1000 less than the average of the entire group. One man in four received increments for duties other than classroom teaching or for being the head of a family. Nearly all of the men elementary school teachers believed that their salaries were inadequate and a majority of them worked outside of school hours to provide the extra income needed. In addition, over two-fifths of the wives were employed, a majority of whom worked of necessity.

The men elementary school teachers who participated in this study recommended a minimum annual salary of $4489 for beginning teachers having the Bachelor's degree and $8071 for teachers with twenty years of experience and the Master's degree. It was interesting to find that a majority of the men favored a single salary schedule for both men and women teachers, meaning that men should not receive additional salary merely on the basis of sex. This viewpoint was also expressed by a large majority of the superintendents, the PTA presidents, and the other educators who participated in this study.

To supplement their teaching salaries, a majority of the men elementary school teachers favored boards of education providing out-of-school work with compensation and over one-half of the superintendents and PTA presidents felt that men would favor such work opportunities. Merit raises, as a plan for increasing salaries, were approved by slightly less than one-third of the men; opposed by nearly one-half; and one-fourth stated that they had not formed an opinion. Many of those who were opposed to merit raises or were undecided stated they would favor merit raises if a workable plan were devised. Others who did not commit themselves said they were not well enough informed on the issues involved to form an opinion. It appeared that a sizable minority of men elementary school teachers did not agree with the National Education Association in condemning merit raises.

A majority of the men chose to teach in elementary schools for positive, professional reasons. Slightly less than one-third had entered elementary school teaching for negative reasons. This latter group represented the majority of those dissatisfied men who, if given a second opportunity, would not choose elementary education. Moreover, they would not encourage other men to be elementary school teachers. In spite of their dissatisfaction and regrets, nearly all of them indicated that they enjoyed teaching in the elementary school.

Although three-fourths of the men teachers and approximately one-half of the superintendents, PTA presidents, and other educators agreed with the statement that men preferred working with men principals, nearly all of the men who had women principals or supervisors said that the women accepted them the same as or better than they did other staff members. Moreover, having more men administrators as a means of encouraging men to enter elementary education was placed seventh by superintendents in their list of recruitment suggestions, twelfth by men elementary school teachers, fifth by PTA presidents, and was not mentioned by the other educators.

Louis Kaplan, in a study made in 1947, asked men teachers whether they preferred male or female administrators. Two-thirds of those responding preferred male administrators and Kaplan concluded that a hostility must exist between men teachers and women administrators.[1] In the present study slightly over three-fourths of the men elementary school teachers indicated that they preferred men administrators. Nine-tenths of those men who had had women administrators, however, said they had been accepted by them. It must be concluded then that men teachers prefer men administrators but not because of animosity shown toward them by women administrators.

---

[1] Louis Kaplan, "The Status and Function of Men Teachers in Urban Elementary schools" (Unpublished Ed. D. dissertation, University of Southern California, Los Angeles, 1947), 367.

Not only did almost all of the men elementary school teachers believe that parents and children accepted them, but approximately one-half said that some parents had specifically requested that their children be assigned to them. Very few men teachers reported that parents had objected to their children being assigned to them.

The future plans of nearly one-half of the men indluded being promoted to elementary school principalships while approximately one-fifth planned to continue being classroom teachers. An additional one-fifth planned to enter related fields in which they had specialized. While one-eighth of the men looked forward to leaving the field of education entirely, the majority of those leaving were not qualified either by desire or preparation to be teachers of elementary school children. If concern were to be felt about these men, it should not be that they were planning to leave, but that they were teaching in elementary classrooms.

Concern should be expressed, however, about the finding that less than one-fifth of the men planned to remain classroom teachers. Although all of the men who expressed the desire to become elementary school principals probably would not be promoted, there would still be enough promotions to constitute a major turnover of men elementary school teachers. The suggestions the participants made to encourage men to remain in elementary education held no answer to the problem of frequent turnover. Although higher salary and greater prestige were ranked high, so were promotions. If promotions were mentioned only as a means of increasing salaries, this could be remedied by higher base salaries. If they were suggested to give men new titles and thus more prestige, the suggestion would serve to compound rather than solve the turnover problem.

Two means of encouraging men to remain classroom teachers were to help them see that men, as elementary classroom teachers, need not be apologetic or fear being stigmatized and that they have a definite role to play and important contributions to make to elementary school children. Contributions frequently mentioned were that men teachers could supplement the feminine influence, that they could adapt the school program to meet the needs of boys experiencing difficulty in school, that they could provide a masculine model, and that they could serve as a father substitute.

If the anticipated turnover of men teachers occurs, then more men will be needed to enter elementary education. The major recommendations that were made to meet this need included better salaries, more active recruitment, and more publicity that would not only be designed to raise the status of all teachers but that would also specifically point out the need for, the contributions of, and the success of men elementary school teachers.

Recommendations

The following recommendations were made as a result of the present study:

1. Men should not be restricted to teaching a particular grade or grades but should be assigned to teach the children they are qualified to teach according to their preparation.
2. To encourage more men to remain classroom teachers, plans should be made to provide adequate salaries, to reduce the concept that elementary school teaching is feminine occupation, and to increase the professional status of all teachers.
3. Boards of education should study carefully the necessity for providing out-of-school work with compensation.
4. Applicants for teaching positions at the elementary school level should be carefully screened to insure that they are properly qualified, in aptitude, interest, and professional preparation, to teach children.
5. There should be no administrative policy that would restrict the number of men teachers assigned to an elementary school.
6. Superintendents should become aware of the opinions held by members of their faculties, including the men teachers; by parents; and by educational leaders regarding the status and role of men teachers in the elementary school.
7. Teachers should be well informed on the issues of merit raises and they should inform their education association concerning their views.
8. High school and college personnel should be more active in selecting and encouraging young men to enter elementary education.

Suggestions for Additional Research

This study suggests a number of related problems which may be investigated.

Many of the men teachers indicated that they had to seek out-of-school employment to supplement their salaries. This raises the question as to what effect such out-of-school employment has on the teaching efficiency of a teacher.

A second area which suggests further study is the sectional differences in salaries. Such a study could determine whether the fact that salaries vary in different sections of the country means teachers are underpaid in some sections or whether other factors are involved, such as variation in the cost of living.

The reasons why men teachers prefer men administrators should be examined to determine whether or not the sex of the administrator has bearing on men remaining in or leaving the teaching profession.

⁂

CORRECTION

Authors Neal Edmund and Lyle H. Hemink wish to correct an error in computation in Table II, page 357, of the May 1960 JOURNAL OF EDUCATIONAL RESEARCH—

> The mean percentage of agreement under "Areas of Least Success" should be 28.10 instead of 21.11.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

USE OF RESULTS OF LARGE-SCALE TESTING PROGRAMS IN INSTRUCTION AND GUIDANCE

ARTHUR E. TRAXLER
Educational Records Bureau
New York, New York

Kinds of Large-Scale Testing Programs

At the outset, let us indicate what we mean by the term "large-scale testing programs." Three types of large-scale testing programs may be identified.

First, there are the programs in which the schools of a particular geographical region may cooperate. These are usually organized on a statewide basis, such as the Iowa State-Wide Testing Program, or on a county basis, as, for example, the testing program in Pinellas County, Florida.

The first state-wide testing programs were begun about forty years ago and have gradually increased in number and size since that time. A survey in 1954 indicated that organized testing programs were carried on in about half the states, either by the state department of education or by one of the state colleges or universities.[1] Currently, Title V of the National Defense Education Act of 1958 is having great influence on the development of state-wide testing programs, and it may be anticipated that in the near future all, or nearly all, states will have such programs.

A second type of large-scale testing program is one in which there is nation-wide participation, either among schools of a particular kind or among all kinds of schools for a special purpose. The annual testing programs of the independent school members of the Educational Records Bureau represent the first kind, and the National Merit Scholarship Examination is illustrative of the latter kind.

In the third place, many of the larger cities in the United States carry on testing programs of their own which are large scale in the sense that they involve all the schools in the city and include thousands of pupils. In 1953, a survey of testing programs in the public schools of cities over 200,000 in population indicated that these programs had a large number of common elements but differed in many details.[2]

In this paper, somewhat more attention is given to the first two kinds of large-scale testing programs than to the third kind.

Organization and Operation of a Large-Scale Testing Program

The opportunities for effective use of the results of a large-scale testing program are determined in part by the way in which such a program is organized and how it is carried on. Provision must be made for the wise choice of tests to be used in a program common to all participating schools; for detailed announcements to the schools listing the recommended tests, indicating testing dates, and setting forth other details; for adjustments to suit the variations and peculiar needs of different schools; for the furnishing of test materials to participating schools; and preferably for services in the processing of the tests after the schools have administered them, although not all wide-scale testing programs make specific provision for the last function.

The nature of a large-scale testing program which is conducted on a national basis may be illustrated by reference to the testing program of the Educational Records Bureau. The Bureau has three kinds of member institutions--independent or private schools, public schools, and colleges. The independent schools, which account for more than 80 per cent of the members, carry on fall and spring testing programs each year. A considerable number of member public schools, particularly those in suburban areas, use the same tests. The tests for these programs are chosen by a committee composed of representatives of member schools--the Test Selection Subcommittee of the Committee on Tests and Measurements. This subcommittee has evolved a rather stable pattern for the Bureau's annual testing programs. Scholastic aptitude tests, tests of reading and spelling achievement, and interest inventories form the main basis of testing in the fall. Achievement tests in nearly all elementary and secondary school subjects are the basis of the spring testing program.

---

[1] "The Status of Statewide Testing Programs," 1954 Achievement Testing Program in Independent Schools and Supplementary Studies, pp. 86-92. Educational Records Bulletin No. 63. New York: Educational Records Bureau, July, 1954.

[2] "The Status of Measurement and Appraisal Programs in Large City School Systems --- A Questionnaire Survey," 1953 Achievement Testing Program in Independent Schools and Supplementary Studies, pp. 75-86. Educational Records Bureau, July, 1953.

Although the Bureau does some test construction, most of the tests used in the ERB program are not constructed by the Bureau but are chosen from those available from the regular test publishers. The tests on the recommended list for each program are announced to the schools, and an order service is maintained for supplying the schools with the recommended tests. Member schools obtain the tests from the Bureau, administer them during a certain period, and return them for processing. The Bureau scores the tests, does the necessary statistical work, and returns an organized report of results together with a letter of interpretation to each school.

Independent school norms are prepared on the basis of the results of tests scored at the Bureau, and percentiles based on these norms are used in reports to the schools. These norms are essentially norms for college preparatory pupils. Some public schools make use of these norms in the interpretation of the test results of their college-bound pupils.

In addition, a school may arrange with the Bureau staff to have its test results recorded on individual record cards and to have these cards kept up to date from one program to another. The staff also carries on research studies which are published in bulletins summarizing the results of the fall and spring testing programs.

The activities of the Educational Records Bureau in the field of testing are in many ways comparable to those that must be carried on in connection with a state-wide or county-wide testing program or by the measurement and research department of a city school system, if it does an adequate job of organizing and reporting test data for use by its teachers and counselors.

Some Prerequisites to Effective Test Use

The foregoing illustration of procedures in a wide-scale testing program points up some of the prerequisites to effective use of test results. One of these is a regular, systematic testing program carried on annually or semi-annually and based on tests which are similar in structure and coverage from one year to the next. Haphazard testing, with the tests chosen willy-nilly without reference to a long-term plan, is next to worthless.

A second prerequisite is development and application of norms appropriate to the group parcicipating in the testing program. It has been found, for example, in the ERB program that the regular publishers' norms are not of much value in the interpretation of the test results of college-bound pupils. Special norms based on results for a large group of college preparatory pupils are needed. Similarly, in certain geographical regions, norms developed in state-wide testing programs are more appropriate than national norms.

A third prerequisite is rapid and accurate processing of tests after they have been administered. In too many schools, teachers and counselors are still expected to carry the burden of scoring and organizing the test results. Almost inevitably, this process is done too slowly by the school staff, and very often it is done inaccurately as well. Efficiency and accuracy can be greatly improved if advantage is taken of modern methods of machine processing, particularly where large public school groups are involved.

A fourth prerequisite is rapid reporting of test results to teachers and counselors in easily used, understandable form. In the ERB testing program, the basic report consists of alphabetical class lists showing the scores and the corresponding percentile ratings accompanied by distributions of the scores of class groups on which the medians and interquartile ranges of the scores are also shown.

A fifth prerequisite is the systematic recording of test results on well-organized cumulative record cards and the keeping of these cards up to date so that growth as well as status of each pupil may readily be apprehended by a teacher or counselor.

Instructional and Guidance Uses of Tests

Except for the fact that large-scale testing programs make possible the development of different kinds of norms, the instructional and guidance uses of tests employed in large-scale programs are not essentially different from those based on tests used in small or medium-sized programs.

It would be unrealistic to try to make a clear distinction between instructional uses and guidance uses of tests. For the most part, the same tests are employed for purposes of appraisal in both these fields. Also, if instruction and guidance are broadly defined, teachers and counselors have similar purposes in the use of the results, and they ought to work in close cooperation.

One use of the test results--and in point of time this is often the first use after a school launches a testing program--is to study the intelligence or scholastic aptitude of the pupils in the local school as compared with that of schools in the country as a whole, or schools in a particular region, or schools of a certain kind. Information concerning the average mental ability in the school and concerning the homogeneity or heterogeneity of the pupils in intelligence is one of the starting points in planning suitable curriculum, in adjusting methods of teaching to ability levels, and in guidance.

A second use of tests is to furnish faculty members with information about the achievement of the pupils in the school in comparison with achievement that would be predicted on the basis of their scholastic aptitude scores. It is important to know whether the pupils as groups are achieving in school work according to their capacity and also to have information as to which pupils are doing better than would be predicted from their aptitude measures and which ones are doing less well than expected. Tests of basic skills and tests of knowledge and understanding in the content areas yield helpful information on these questions. At the same time, it is necessary to keep in mind the fact that there is no sharp differentiation between scholastic aptitude tests and achievement tests and also to remember that the objective tests used may not always be well suited to the school's own courses.

A third important use--an instructional use--is to help classroom teachers know the level of ability and achievement of the classes assigned to them. At the beginning of the term, the teacher may study the scholastic aptitude records of the pupils in his classes and also the achievement test records of the pupils; thus he may plan his instruction at a level suitable for the pupils. The level and kind of instruction needs to be different for a quick-to-learn, fast-moving group than for a slower-learning group that may require a good deal of reteaching. Similarly, the teacher needs to plan beforehand for differentiation of instruction to take account of varying abilities and of needs for special attention. With test data at hand, the teacher can be better prepared to begin work with his classes, and tests administered during the course of instruction or in the spring near the end of the school year will help him see how effective his instruction has been and how much the pupils have developed.

A fourth use is to help teachers and counselors study the achievement of individual pupils in comparison with the scholastic aptitude of these pupils. How is each pupil progressing in each field of study in comparison with his capacity? It may be found that a pupil whose progress is slow in comparison with his class group is doing very well indeed when his achievement test percentiles are compared with his scholastic aptitude test percentiles, while a more able pupil who is easily keeping up with his class may be working far below his potential. These kinds of information have marked value for both the individualization of teaching and individual counseling.

A fifth use of test results--one of the most fruitful of all--is to furnish teachers and counselors with an objective basis for the diagnosis of strengths and weaknesses of class groups and particularly of individual pupils. For instance, a pupil's achievement profile on a test battery such as the Metropolitan Achievement Tests may be studied, and steps taken cooperatively by his teacher and counselor to help him take advantage of his strong points and to plan special help or remedial work in order to bring up the low points in the profile. Usually, of course, diagnosis of weaknesses calls for more detailed techniques than are found in a test of general achievement, but if broad areas of strengths and weaknesses can be identified by means of scholastic aptitude and achievement tests administered to all pupils at a given level, then more specific diagnostic tests may be used with those pupils found to be low on the survey test.

A sixth use of test results is to aid in the discovery of special abilities of individual pupils which should be developed. For example, annual participation in a large-scale testing program and systematic recording of results on cumulative records may point up outstanding abilities of individuals in certain fields, such as science or foreign languages. This kind of use of tests may be extended through supplementary administration of tests in special fields, such as art or music, which ordinarily are not a regular part of programs of large-scale testing.

A seventh use of tests is to help teachers and counselors gain better insight into the interests of pupils. Clues to a pupil's interests may, of course, be obtained by observing and talking with him and noticing what he reads and what his leisure time activities are. Informal procedures of this kind are necessarily used with elementary school pupils, but, beginning with the junior high school and especially in the senior high school, standardized inventories of interests may well become a regular part of wide-scale testing programs. Interest inventories should not be used alone, but, when they are employed in conjunction with scholastic aptitude and achievement tests and tests of specific abilities, they frequently contribute to the school's understanding of the pupil.

An eighth use of tests by schools is, of course, in the guidance of individual pupils toward educational and vocational goals. This use is related to most of the other uses which have been mentioned. It involves careful individual study and counseling of pupils at all ability levels. In this connection, the most important instrument of all is the cumulative record. If a series of comparable tests covering the different fields of study is used annually and if the results are brought together in organized form, either numerically or graphically, one can compare the individual's scores from year to year in the different fields and draw conclusions concerning his level of ability and his progress and can help him and his parents to see his potential for further education and to adjust vocational aspirations in accordance with the indicated facts.

A related use of test results is in counseling having to do with problems of adjustment. Scores on tests of scholastic aptitude, special aptitudes, achievement, and interests often provide useful leads in identifying symptoms of maladjustment in the school situation. It is well known, for instance, that there is frequently a relationship between marked retardation in reading and lack of adjustment to the group. Thus, reading test scores, and, to a lesser degree, scores in arithmetic, spelling, and other basic tools furnish important clues to adjustment needs.

In certain situations, personality inventories may also be used in a limited and experimental way in advising pupils concerning adjustment problems, although this kind of measure is seldom used regularly in wide-scale testing programs. The question of whether or not a school should supplement the results of tests in a wide-scale program with personality tests needs to be decided in the light of the availability of trained personnel for competent interpretation of the results. Informal methods of personality appraisal usually are to be preferred, except in schools that have access to the services of psychologists.

In the guidance use of the results of tests, assistance to the individual in meeting specific educational, vocational, and adjustment problems is important. But the central, long-term purpose is to help the individual reach greater self-understanding, self-appraisal, and self-dependence. Hence, the individual student as he matures should gradually be helped to understand and interpret his own test results and to see their relationship to other kinds of information about him.

The last use of the results of wide-scale testing programs to be mentioned here is in long-term research on a wide variety of problems pertinent to the education, guidance, and development of individuals and groups. One of the best current illustrations of this kind of use is the projected twenty-year Talent Study directed by John C. Flanagan of the American Institute for Research, with financial support from the Office of Education and other agencies of the Federal Government.

Pre-service and In-service Education in Testing

At the present stage of educational measurement in the schools of the United States, there are still problems with regard to the kind of instruments to prepare and problems of test construction that have not yet been met, or at least have not been met well. But the major problem is no longer one of technique of test construction, if it ever was. It is a problem of pre-service and in-service education of test users. Not only should counselors be well trained in measurement, but every teacher should have at least one elementary course in tests and measurements, followed by a continuous in-service training program. The central question is one of communication. How are schools to communicate their needs to test specialists, and how are the specialists to communicate their knowledge of the values and limitations, the uses and misuses of instruments of appraisal to users?

There is no easy answer, for testing, even when reduced to the simplest possible level, is still somewhat technical. The assessment of human abilities, achievements, and personal qualities is inevitably a complex business, and it is not possible to convey its essentials to school staff members in one easy lesson. It would seem that the only way of meeting the need is through a continuous program of in-service education, including selected reading materials, study and discussion sessions, illustrated lectures and discussions, case conferences carried on by groups, and numerous individual conferences. The value of test results to a school is likely to be directly proportional to the imagination, resourcefulness, and vigor with which the in-service program is planned and developed.

Summary

In summary, I have tried to say that there are three types of large-scale testing programs, including those organized on a geographical basis, those organized on a national basis either for schools of a particular kind or for a specific purpose, and those which are city-wide; that the value of the results of large-scale testing programs is partly determined by the organization and procedures of carrying on these programs; that among the prerequisites to effective use of test results are a regular, systematic testing program, the development and application of appropriate norms, rapid and accurate test processing, rapid reporting of test results in a simple, understandable form, and systematic recording of test data on individual cumulative record cards; that the uses of the results of wide-scale testing programs include (1) assessment of the general level of intelligence or scholastic aptitude in the school, (2) comparison of the general level of achievement in the school with the general level of scholastic aptitude, (3) use by classroom teachers in studying the level of ability and achievement of their class groups, (4) use by teachers and counselors in studying the achievement of individual pupils as compared with their scholastic aptitude, (5) diagnosis of the strengths and weaknesses of class groups and individual pupils, (6) discovery of special abilities of individual pupils which should be developed, (7) identification of the major interests of pupils, (8) guidance toward educational and vocational goals, (9) counseling with problems of adjustment, and (10) extended research on educational and guidance problems; and, finally, that the major problem in the use of the results of wide-scale testing programs is improvement of pre-service and in-service education of test users.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

THE ASSERTION OF AUTHORITY: DIFFERENCES BETWEEN TEACHERS, STUDENT-TEACHERS, AND MOTHERS OF YOUNG CHILDREN

ALBERT S. DREYER and DOROTHY HAUPT
The Merrill-Palmer School
Detroit, Michigan

To what degree do parents and teachers assert the power and authority vested in the roles assigned them? To what degree is there similarity between the power asserted by mothers and that asserted by teachers, and in what areas of child rearing? Do students during their period of supervised teaching perceive their authority role to be similar to that of the regular teacher? Do they assert the authority vested in the role they are soon to adopt?

As part of a larger study concerned with the effect of nursery school experience on children's kindergarten behavior, the authors investigated the extent to which the principal adults in the child's life, namely, his mother and teachers, express the power and control vested in their respective roles. The present study was designed to determine the differences in the assertion of power between the mothers of nursery school children and the teachers and student-teachers of the children.

Procedure

An instrument used by Harrison (1), consisting of 11 brief story completion situations intended to show adult reaction to children's behavior in various situations, was adapted to the study. This questionnaire, in which the adult is asked to indicate how he would handle each situation described, was given to 38 middle-class mothers of nursery school children in the Detroit area, 7 kindergarten teachers whose classes these children would enter, and 11 experienced teachers and 9 student-teachers of the children at the time of testing. The total of 65 questionnaires was then coded for the degree of power assertion expressed.

In the coding, each story was rated on a 5-point scale of power-assertive behavior, defined as the extent to which the adult exerted direct, coercive pressure on the child intended to change his behavior immediately. The adult's behavior was further considered power assertive to the degree that it was arbitrary and allowed the child no choice, and therefore no compensatory gratification. Examples of the characteristics of power-assertive behavior used in scoring follow:

Pressure

Choice-giving vs. no choice-giving
Coercive vs. noncoercive regarding compliance

Examples:

High pressure: "You must drink your milk."
Low pressure: "Would you like some milk?"

Arbitrariness

Adult's desires vs. other considerations primary

Examples:

High arbitrary: "Don't throw that shoe because I don't want you to."
Low arbitrary: Don't throw that shoe because you'll hurt Jimmy." "That window will break if you hammer it."

Immediacy

Immediate compliance or no closure vs. some closure allowed regarding compliance

Examples:

High immediacy: "Stop that right now."
"Stop!"
Low immediacy: "Come in after you finish that puzzle." "Come in the house in five minutes."

TABLE I

POWER ASSERTION SCORES OF SUBJECTS

	Mothers N=38	Kindergarten Teachers N=7	Nursery School Teachers N=11	Student Teachers N=9
Score: Mean	3.00	3.46	3.20	2.29
Variance	0.61	0.86	0.37	0.09
		H test = 11.26		
		d.f. = 3		
		P < = .02		

TABLE II

MEAN POWER ASSERTION SCORES OF SUBJECTS ON SHARING AND ACHIEVEMENT SITUATIONS

	Mothers N=38	Kindergarten Teachers N=7	Nursery School Teachers N=11	Student Teachers N=9
Mean PA Score				
Sharing (Story 3)	3.45	4.00	4.00	2.89
Achievement (Story 5)	2.60	3.74	3.91	2.67
Wilcoxon sign rank test	2.34	.....	.....	.....
Level of significance	< .02			

A sample of the stories of eight subjects, not finally used in the analysis, was coded by three judges, the two authors and and a graduate student in child development. Coefficients of concordance (2) were calculated for the ratings, to determine which stories could be reliably coded. For 5 of the 11 stories the coefficients were significant at the 5 per cent level or better; they ranged from .72 to .81. The final power assertion score used was the mean score for each subject on these 5 stories. The coefficient of this mean power assertion score was .91, significant at the 1 per cent level.

Stories Used in the Study

The following five story completions were used:

1. Some people expect children to show adult table manners when they are around four years old. Others feel this is an area which should be left alone until later. For example, how would you usually handle it if your child spilled his juice on the table?
2. "Read to me" or "Read me another story" appears to be a favorite request of most young children. There are times when they will listen to story after story. At other times they will listen for a time and then begin to talk about things unrelated to the story you are reading. What do you usually do when your child begins talking when you are trying to read to him?
3. Some people feel it is very important to teach the preschool aged child about sharing. Others give this little attention at this age. Let's take this as a situation. Your child is playing with all the cars in the room. Some other child comes over and asks him for a car. He refuses. What would you usually do in such a situation?
4. Children have a tendency to get over excited at times, having a marvelous time but shouting quite loud. If, at such a time, you were to remark to your child that he was having a good time, but wouldn't he try to be a bit quieter, and he didn't respond but kept shouting, what would you do? If he still didn't stop, what would you do?
5. Many people have raised questions about when to help young children solve their problems and how much help to give them. For example, your child is trying to make a "snowman" from clay. He tries to put arms on it but each time they fall off when he lifts it. What would you do in a situation like this?

Results

The Kruskal-Wallis H test (3), a nonparametric analysis of variance, was calculated for the four groups. The test was significant at better than the 2 per cent level, indicating differences in mean power assertion scores; that is, the student-teachers had lower scores than the mothers or teachers. (Table I). Separate tests of significance, using the White test (3) were calculated, again showing the student-teachers to differ significantly in mean power assertion from the other groups, while the mothers and teachers did not differ significantly. F tests (3) as well indicate that the student-teachers are not only on the average less power assertive but also are more homogeneous than the other groups; that is, their variance estimate is significantly smaller than that of the other groups.

Since the framework of this study involves the role demands of teacher and mother, an attempt was made to determine how far the problem of child control is general to both. Accordingly, separate mean power assertive scores were determined for two stories that seemed salient for both roles, namely, story 3, involving a sharing situation, and story 5, involving an achievement-related situation. Table II shows that the mothers significantly differed in power assertive scores as between the sharing and achievement-related stories. The difference amounted to a drop of almost 1 scale point. Although there was a slight drop between the two stories for both teachers and student-teachers, it was not statistically significant. For the mothers the particular conditions expressed in the stories appear to stimulate the assertion of power far more than for either group of teachers.

Some further support for this interpretation is provided by an analysis of the differences in power assertive behavior between seven nursery school teachers who were wives and mothers and three who were not, as compared to the group of mothers. (Information about the marital status of one teacher was lacking.) No significant differences were found in the mean power assertion scores between teachers who had children and the unmarried teachers, but there was a tendency (White test significant at the .08 level) for the teachers with children of their own to be somewhat more power assertive in their responses than were the group of mothers. In the absence of any evidence that nursery school teachers are more power oriented than other groups of women, it would appear that the role demands of teacher are different from the role demands of mother with respect to expressing power. A mother in the role of teacher seems to be more assertive in expressing power than are mothers not in this position.

Discussion

The present study indicates that of the three groups the student-teachers are most inclined to allow the child choice and least inclined to assert power or authority in response to the adult-child story completions. The sheer fact of being an adult may give the student-teacher authority in the eyes of children, but any expression of such authority is under constant supervision and evaluation. In addition, the supervising teacher's perception of her role with students may be so inclusive that the student-teacher as a learner is accorded little opportunity for exploring the behavior suitable to various roles. Thus the difference in role of the person who has actual control and responsibility for maintaining control probably accounts for this finding.

It may be possible that student-teachers have an essentially different philosophy of education from that of experienced teachers. However, from observing and interviewing them for other reasons we gained the impression that this was not so. Even were it so, it would not explain their differences from the group of mothers. The difference is more plausibly explained by their different role, as already noted, and their lesser amount of experience in dealing with children. However, the factor of differences in concepts of education should be controlled in further research.

Certain personality factors, such as the value system or general orientation of the teacher toward authority and power, might also be studied and controlled in future research. For example, do students who express feelings of insecurity in relation to the control of children, before they have direct experience with them in a group, exert tighter control when their authority is threatened? Do students who initially express confidence and comfort in relation to asserting control or power convey this feeling to children in actual contact with them? The degree to which the student expresses power as a student-teacher may be predictive of whether or not she will make an able or competent teacher in terms of the managerial aspects of the role. That is, the degree to which the student's personality is consonant with this important requirement of the teacher's role provides a good predictive factor.

The findings on changes in power assertive behavior from story to story suggest that the controls demanded of the teacher are less subject to situational factors than are those of the mother. As compared with those of the mother, the teacher's control problems are multiplied by the sheer number of children involved. As a result, the teacher may be less sensitive to the contexts or motivations of the individual child's behavior. Further, the child's mother, because of her emotional involvement, and thus greater sensitivity to the child's motivations, may be able to discriminate differences in the child's behavior from one situation to another. Again, if the child does not share with another, the mother may feel that her ability as mother is called into question. Under the same circumstances, the teacher's self-esteem is not threatened. Again, the teacher's knowledge of the children's readiness to share and her expectation that they will do so are reflected in her responses. An intensive analysis of such facets of the teacher's role with a view to seeing how role demands and personality needs are congruent and lead to satisfactions with the teaching profession, would be of value.

Summary

A set of five story completion situations, depicting adult-child interaction, was presented to 38 mothers of nursery school children, 11 nursery school teachers, 9 student-teachers in the nursery school, and 7 kindergarten teachers. Each was asked to indicate how she would respond to the situation described. Each story was scored for the degree of power assertion expressed, and a mean score was obtained for each person. Analysis of the data indicated that the four groups differed significantly in the mean power assertion expressed. The student-teachers differed significantly from the experienced teachers and the group of mothers. The experienced teachers and mothers did not differ significantly from each other. Data bearing on the similarities and differences in the control demands of teacher and mother are presented.

REFERENCES

1. Harrison, H. K., "An Investigation of Transfer from Children's Expectations of Mothers to Children's Expectations of Teachers," Psychological Newsletter, 1956, 7, 107-120.
2. Rogers, D., Mental Hygiene in Elementary Education. Boston: Houghton, Mifflin, 1957.
3. Edwards, A. L., Statistical Methods for the Behavioral Sciences. New York: Rinehart, 1954.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

A SURVEY OF EDUCATIONAL DISABILITY IN EMOTIONALLY DISTURBED CHILDREN*

ARTHUR S. TAMKIN
Children's Psychiatric Hospital
Columbus, Ohio

Research concerning the processes underlying educational disability in children has demonstrated a complex set of relationships between emotional problems and educational achievement. While most investigators concur in the general finding of a close coincidence of emotional problems and learning disabilities, they often vary with respect to the theories they advance to account for this fact. Blanchard (1), after evaluating the relations between reading disability and difficulties of personality and emotional development in 73 consecutive cases referred to the Philadelphia Child Guidance Clinic, concluded that there was no causal relationship. She offered the theory that "the reading disability often arises from the same source of difficulty in emotional development, and in much the same manner as the accompanying personality or behavior problems or neurotic symptoms..." Support for postulating a causal relationship may be derived from the studies by Fendrick and Bond (3), who found a strong relationship between juvenile delinquency and reading disability, and by Gates and Bond (4), who observed a wide variety of symptomatic reactions attributable to reading disability. The latter evaluated 100 reading disability cases and found that out of this group approximately three-fourths showed personality maladjustment, of which 25 percent were considered to have personality problems causing the reading disability and 75 percent were considered to have personality problems as a result of the disability. Their findings thus permit the inference that by and large a learning disability causes emotional and personal maladjustment. It is often pointed out by proponents of this view that a learning disability disrupts the child's functioning in the school situation, making for frustration and difficulties in an important sphere of social adjustment. Severe reading disability, it is maintained, is likely to be a serious catastrophe to many children who feel defeated in a very significant aspect of living. And finally, as Brueckner and Bond (2) point out, there are studies which show that some children with severe emotional problems compensate for their difficulties by excessive educational striving, thus becoming extremely competent students.

In addition to the role of emotional and personality problems in educational disability, another important factor is what might be termed an unfortunate home condition. Siegel (6), in his study of personality structure found in reading disability, has called attention to the importance of the home and family relationships to the educational adjustment of the child. Clearly, the significance of the home lies in the fact that family relationships condition the child's ability to follow directions, his use of language, and his background of understanding. Modes of family life which interfere with the orderly development of these basic abilities thereby contribute to educational failings in the school situation.

On the basis of research findings such as these, which demonstrate complex relationships between emotional adjustment, home conditions, and educational disability, one would expect to find rather severe educational disability in children with emotional disturbances requiring residential psychiatric treatment. In this connection Jastak (5) observed that high reading but low arithmetic scores tend to occur in abnormal mental states of a developmental nature of long standing as in neurosis and schizophrenia in both adults and children.

The purpose of the study reported here was to survey the educational achievement of emotionally disturbed children to ascertain the frequency of educational disability and to determine the relationship between arithmetic and reading skills. It was hypothesized that in a group of hospitalized, emotionally disturbed children educational disabilities would predominate and that arithmetic would be lower than reading ability.

---

* This study was supported by the Ohio Division of Mental Hygiene. The opinions expressed herein are those of the author alone and not necessarily those of the Division of Mental Hygiene.

Method

The Ss consisted of children receiving residential treatment for emotional disorders at a psychiatric hospital. Shortly after admission they were given the Arithmetic and Reading parts of the Wide Range Achievement Test by the teachers of the hospital school.[1] They were selected for testing on a random basis, and their obtained scores were taken as measures of educational achievement in reading and arithmetic. In all there were 22 boys and 12 girls with mean ages of 8.7 and 9.4 respectively. The combined mean age was 9.0 years. A special characteristic of this sample and of the population of patients under consideration is that they are devoid of patients with serious medical problems, with psychiatric symptoms requiring isolation or confinement, or with mental deficiency. The absence of mental deficiency was determined either by an IQ measure above 70 or by clinical estimation of the admitting professional staff. Since IQ's were not available for all Ss. chronological age was used as the basis for evaluating educational disabilities in these cases.

Results

The mean grade rating on the Wide Range Achievement Test was 3.48 for the entire sample. For reading alone it was 3.81 and for arithmetic it was 3.14. The difference between grade ratings for reading and arithmetic is significant at the .005 point based upon a one-tailed test ($t = 2.91$).

In order to evaluate educational disability, the achievement grade ratings were subtracted from the chronological ages of all Ss. Remainders lying between 5.0 and 5.9 would indicate a level of achievement commensurate with the chronological age since a child normally enters the first grade of school between the ages of 5 and 5.9 years. Accordingly, remainders above 5.9 would indicate the presence of disability, while those below 5.0 would indicate advanced achievement. The mean difference between chronological age and grade rating for the sample was found to be 5.44, suggesting that the achievement of these emotionally disturbed youngsters as a whole is commensurate with their chronological age. The mean difference for the boys was 5.28 and 5.73 for the girls. The difference between the sexes is not significant at the .05 level, however.

Observing the difference between chronological age and grade rating for each S, it was noted that 32% of the cases (n = 11) demonstrated some degree of educational disability. However, 41% of the cases (n = 14) were educationally advanced, and the remaining 27% (n = 9) were at their expected grade level.

Discussion

The expectation that educational disabilities would predominate in a group of emotionally disturbed children requiring residential treatment was not borne out by the findings of this study. The Wide Range Achievement Test scores showed that for the group as a whole the academic achievement level was consistent with the expectation for its mean chronological age. Furthermore, an analysis of each individual's achievement level showed that the bulk of the Ss were either performing at expected grade level or even educationally advanced relative to their chronological ages. Only one-third of the Ss showed signs of some learning disability. Assuming that the Wide Range Achievement Test produces a valid measure of achievement, the findings are most compatible with the theory that educational disability is merely a symptom which might be produced by the same condition responsible for the emotional disorder. Certainly, emotional problems could not be construed as a cause of educational disability since, as the results demonstrate, only one-third of the emotionally disturbed sample had accompanying educational disabilities.

A possible weakness of this study which could affect the results might be the use of chronological age rather than mental age as the basis from which to measure departures in educational achievement. This choice was necessitated by the unavailability of IQ's for all cases in the sample. The justification for this procedure was predicated upon the assumption that the Ss would not show extreme dispersion about the average IQ range.

The hypothesis that emotionally disturbed children would score higher on the Reading part than on the Arithmetic part was confirmed. Since the mean grade ratings of both parts fell between 3.0 and 4.0, however, it may not be inferred that emotionally disturbed children tend to present arithmetic disabilities since for a group with a mean age of 9.0 years grade ratings within the obtained range normally would be expected.

Summary and Conclusions

Two hypotheses were presented and tested by means of a survey of the educational achievement of emotionally disturbed children. The hypothesis stated that educational disabilities would predominate and that arithmetic would be lower than reading ability. Thirty-four children recently admitted for residential treatment at a psychiatric hospital were used as Ss, and the Arithmetic and Reading parts of the Wide Range Achievement Test were used to assess educational achievement. Failure to confirm the first hypothesis lent support to the theory that educational disability may be a symptom of the same condition producing the emotional problems rather than being caused by the emotional problems. The second hypothesis was supported by the data, although both arithmetic and reading grade rating were within the range commensurate with the mean chronological age of the sample.

---

[1] The author is grateful to Miss Jeanne Hugo, School Principal, who made the Wide Range Achievement Test data available for this study.

REFERENCES

1. Blanchard; P. J. Reading disability in relation to difficulties of personality and emotional development. Mental Hygiene, New York, 1936, 20, 384-413.
2. Brueckner, L. J. and Bond, G. L. The diagnosis and treatment of learning difficulties. New York: Appleton - Century - Crofts, 1955.
3. Fendrick, P., and Bond, G. L. Delinquency in reading. Pedagogical Seminary and Journal of Genetic Psychology, 1936, 48, 236-243.
4. Gates, A. I., and Bond, G. L. Failure in reading and social maladjustment. Journal of the National Educational Association. 1936, 25, 205-206.
5. Jastak, J. Wide Range Achievement Test. Manual. Wilmington: Charles L. Story, 1946.
6. Siegel, M. The personality structure of children with reading disabilities as compared with children presenting other clinical problems. Nervous Child, 1954, 10, 409-414.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

EXPERIMENT IN CROSS GROUPING

JEFF WEST and CALLIE SIEVERS
Dade County Public Schools
Miami, Florida

The Problem

A study in cross grouping has been under way for several years at Central Beach Elementary School, Dade County, Florida. It was observed that certain elementary school pupils learned their fundamental linguistic and arithmetical skills in about one half the time it takes the average child to master them. It seemed a waste of time and talent to hold these pupils in a standard classroom situation for the full school day. On the other hand, isolated ability grouping for the total school day presented a number of problems such as the development of undesirable attitudes and relationships between pupils and classes of pupils. There is also the problem of denial of leadership and stimulation which high ability pupils might provide for their classmates. It was decided to test the effectiveness of a compromise type of school organization that would challenge the high ability pupils and would also provide them with common experiences with other pupils for a part of the school day.

Procedures

The high ability pupils were identified by the teachers and through a testing program administered by psychologists from the Diagnostic Center. The term cross grouping is derived from the fact that the fifth and sixth grade teachers send their high ability pupils from their heterogeneously grouped classes to a resource group teacher for the morning session. These pupils then return to their regular classrooms for the afternoon session. The third and fourth grade teachers keep their high ability pupils with their heterogeneously grouped classes in the morning and send these pupils to the resource group teacher in the afternoon. The term cross grouping is applied because the high ability pupils are moving from a classroom situation with pupils of all ability levels into a second learning situation with pupils of similar abilities and because grade levels are interacting with each other in this organizational movement. The purpose of the resource group is to extend the pupils' competency in the basic skill areas through practical application and through creative experiences. Considerable emphasis is placed upon the talented pupils' responsibilities for service to their fellow pupils, parents and teachers. Parental understanding and systematic coordination between the regular classroom teachers and the resource teachers has been developed through the leadership of the principal. The consultant services of the Horace Mann Lincoln Institute of School Experimentation, Columbia University, were utilized in the evaluation of this experiment. The following instruments and procedures were used to measure the results of this program.

Educational Development

Standardized Achievement Test Scores

Metropolitan Achievement Tests
Iowa Study Skills Test

Interest Inventories

Group Composition Questionnaire

Identification Questionnaire

Teacher Judgment

School Records and Reports

Personal Development

Interest Inventories

Iowa Study Skills

Personal evaluation

Teacher evaluation

Social Development

Sociograms of classroom groups and resource group

Group composition questionnaires

Identification questionnaire

Teacher judgment

Record of participation in the school activities

The Interest Inventories, Group Composition Questionnaires, Identification Questionnaires, Guess

Who Questionnaires, were adaptations of the forms presented in Ruth Cunningham's Understanding Group Behavior of Boys and Girls, New York, Bureau of Publications, Teachers College, Columbia University, 1951.

Population

Resource Group	8 sixth grade students (5 boys and 3 girls) 8 fifth grade students (6 boys and 2 girls)
Regular Classrooms	110 sixth grade students 108 fifth grade students
Central group	Students, matched by intelligence score, sex, and chronological age, attending a school having a similar population, located in a similar environment, and in which no special grouping had been practiced

Results

Educational Development

1. The Total Grade Placement Scores indicated that the students in the Resource Groups achieved more than two years above both the national median and their classmates in the regular classroom.

2. The Total Grade Placement Scores indicated that:

   a. The Resource Group (5th grade) achieved five months beyond the Control Group.

   b. The Resource Group (6th grade) achieved two months beyond the Control Group.

3. The Iowa Basic Study Skills Test scores indicated that the Resource Groups achieved more than one year in advance of the regular groups.

4. A study of the Interest Inventories indicated that the difference in the interests of the two groups was in the degree of interests rather than in the area of the interests. However, the pupils in the Resource Groups indicated more interest in science than those pupils who were not included in the Resource Groups.

Personal Development

1. The Resource Groups had fewer problems in the area of human relationships. There was a greater degree of confidence expressed by the Resource pupils in their relationships with other people. The results of the Iowa Basic Skill Tests indicated that the Resource pupils have a more advanced mastery of the study skills which contributed to their superior ability to work independently.

2. Pupil self-evaluations indicated increased ability to work with others, to cooperate, to accept responsibility, to use more and better materials, to do creative writing, and to work independently.

3. There appeared to be fair agreement between the ratings of the classroom teachers and the teacher of the Resource Groups.

4. A comparison of the "Guess Who" ratings indicated that the Resource Groups seemed to possess a greater degree of insight into their own characteristics and behavior patterns than the total group.

5. Sociograms revealed that the majority of the talented students from the Resource Groups fell into a normal pattern of social relationships both in the Resource Group and when they were scheduled with other pupils in the regular classes.

A follow-up study of 28 students from the Resource Classes was made into the junior and senior high school. An analysis of the secondary teacher evaluations concerned with ranking personality characteristics revealed:

1. No student received ratings in all areas in the one percentile range.
2. Thirty-five percent of the students received no rating below the top 25% range.
3. Leadership is not high in the number of ratings in either the top 2% or the top 10%.
4. Leadership is high in the top 25%.
5. Working independently, persistence, responsibility, interest, curiosity, and the ability to think creatively received higher ratings in the top 2% and the top 10% rankings.

The achievement summary for these pupils as tested in the seventh, eighth and ninth grades follows:

10th Graders, Test given in 9th grade		
National Median		9.7
Gifted Medians:	Reading	10.9
	Arithmetic	10.3

9th Graders, Test given in 8th grade		
National Median		8.2
Gifted Medians:	Reading	10.2
	Arithmetic	10.8

8th Graders, Test given in 7th grade		
National Median		8.2
Gifted Medians:	Reading	10.4
	Arithmetic	8.5

7th Graders, Tests given in 7th grade		
National Median		7.2
Gifted Medians:	Reading	10.0
	Arithmetic	8.6

Conclusions

1. No significant loss in social relationships between pupils was evidenced in cross grouping.
2. Significant gains were made in academic achievement.
3. The subjective judgment of teachers indicated that balanced cross grouping provided increased opportunities to develop creative talents and to encourage aesthetic appreciation.
4. It is important that lines of communication be kept open between all persons involved in the program.
5. It is vital that the program be carefully planned in terms of purposes and objectives to insure a balance between language arts, science and mathematics.
6. The identification of high ability pupils in the elementary school should include more than one measure of ability because of their immaturity and the inability to respond accurately to measuring instruments which require reading and writing.
7. The quality of leadership was more frequently found among those pupils ranking in the lowest section of the top quartile than among those in the top 10%.
8. Cross grouping does not automatically achieve the desired goals but can facilitate the achievement of them.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

COUNSELOR AND STUDENT ESTIMATES OF GRADES AS PREDICTORS OF ACADEMIC ACHIEVEMENT

MALCOLM H. ROBERTSON*
University of Florida
Gainesville, Florida

Traditionally, academic achievement in college has been predicted by correlating test scores and/or high school grades with first semester grade point averages. With the rise in popularity of research on the self-concept variable, consideration has been given to the possibility of using student self-estimates and/or counselor estimates as predictors of grades. The rationale for this approach seems to be that in addition to information about test scores, counselors and students can take advantage of relevant but essentially unmeasurable variables in their prediction of grades.

The few studies utilizing student and counselor estimates of grades have yielded interesting but conflicting results. For example, Young (7) obtained a correlation of .71 between student estimates of grades and actual grades. In the same study, the correlation for counselor estimates with grades was .68. On the other hand, Budd (1) found a very high (.76) correlation between counselor estimates and actual grades but a low (.29) correlation between student estimates and grades. Freehill (2) also reported a low (.23) correlation of student estimates with grades earned at the end of the first semester in college. In a study of student aspirations at the beginning of college, Matteson (3) obtained a significant association between predicted and obtained grade point averages. Finally, Robertson (4) compared student estimates with counselor estimates of freshman grades and found a statistically significant agreement between the two sets of predictions.

In view of the promising but inconsistent findings, the present study was undertaken with the hope of providing some definitive conclusions. Students attending a precollege testing and counseling program were asked to predict their first semester grades both before and after the precollege program. Counselors were asked to predict the grades for each student they counseled. The two sets of predictions were then compared with grades received at the end of the fall semester.

Procedure

The precollege counseling and testing program consisted of six three-day sessions conducted during the summer at the University of Mississippi. On the first day of the session, students filled out a routine questionnaire. One item on the questionnaire required the student to indicate the expected grades in five courses for his first semester. No specific courses were mentioned. Next, students took a battery of tests consisting of the Differential Aptitude Test, the Cooperative English Tests, the Strong Vocational Interest Blank, the College Qualification Tests, and the Diagnostic Reading Tests. On the last day of the session, each student had a 45-minute interview with a counselor.

Using percentile ranks, the counselor made a direct interpretation of the test scores and informed the student of the type of educational program indicated by the pattern of his scores. Upon completion of the counseling interview, the student completed another questionnaire, again predicting the grades for five courses for the first semester. On both questionnaires the student used a code number instead of signing his name. The counselors also filled out a questionnaire for each student after the interview. One item of the questionnaire required the counselor to predict the grades that he thought the student would make on five courses for the first semester. The counselor signed his name to the questionnaire. While neither the student nor the counselor was given any special criteria to use in making predictions of grades, both were presumably guided to some extent by the curriculum the student planned to take.

Of the 200 students who participated in the program, approximately 50 per cent intended to enter Liberal Arts. Of the remaining students, 15 per cent planned to enroll in Business School, 15 per cent in Engineering, 10 per cent in Education, and 10 per cent in Pharmacy. Slightly more than half the students were males, and, with the exception of a few older students, the age range was from 17 to 19 years.

Five counselors participated in the study. Two were University staff counselors and one was a University High School counselor. The remaining two were staff members of the Student Personnel Division who had been given an intensive course in the type of counseling used in the program.

---

* Formerly of the University of Mississippi

Results

As described in the preceding section, the data consisted of student predictions of grades before and after the testing and counseling program, counselor predictions of grades, and the grades students actually earned at the end of the first semester. Six separate analyses were made and the results are presented in Table I.

First, student pre-estimates of grades were compared with their post-estimates. A contingency correlation of .55 was obtained and for a df of 1 this value is significant below the .001 level. Since the size of the contingency correlation is limited by the number of categories in the table, the maximum correlation that could be obtained for a 2 x 2 table if the variables were perfectly correlated is .70 rather than the conventional value of 1.00 (5). To give the reader a better idea of the closeness of the association in this first analysis, an index of order association, 0, was computed and a value of .94 was obtained.

Second, when student pre-estimates were compared with obtained grade point averages, a correlation of .38 was found. When student post-estimates were compared with grades, the correlation was .37, and for a df of 1 both correlations are significant below the .001 level. Again, with a 2 x 2 table the maximum correlation that could be obtained if the variables were perfectly correlated is .70.

Next, the correlation between counselor estimates and grades was .50, and for a df of 4 this value is significant below the .001 level. This correlation of .50 was based upon a 3 x 3 table and the maximum value that could be obtained is .81 as compared to the .70 correlation of student estimates with grades.

Finally, both sets of student estimates were compared with counselor estimates. A correlation of .40 was found for the student pre-estimates and counselor estimates, while the correlation between student post-estimates and counselor estimates was .50. For a df of 6, both values are significant below the .001 level. Since both correlations were based on a 4 x 3 table, the maximum correlation that could be obtained falls somewhere between .81 and .86.

Discussion

The findings presented in the previous section can be organized as follows: First, a very close correspondence was found between student estimates of grades made before counseling and those made afterwards. Possibly the testing and counseling program confirmed the student's opinion of his own abilities, or perhaps his opinion was not susceptible to much change. This close correspondence of student pre-estimates with post-estimates was also manifest in the correlation of both these estimates with grades. That is, the student estimates made before counseling correlated as well with later grades as did the estimates made after counseling.

Since student post-estimates correlated higher with counselor estimates than did the student pre-estimates, and since counselor estimates correlated higher with grades than did the student estimates, it might seem surprising that student post-estimates did not correlate higher with grades. However, this higher correlation of counselor estimates with grades is partly explained by the larger (3 x 3) table of frequencies as opposed to the smaller (2 x 2) table of frequencies for the correlation of student estimates with grades.

Though all the correlations were statistically significant, they were only moderately high. As mentioned before, the size of the contingency correlation is limited by the number of categories. In the case of student estimates, so many of the ratings were concentrated in the B and C grade categories that it was necessary to break a larger table down to a smaller 2 x 2 table by combining frequencies in the more extreme grade categories. While this factor might account to some extent for differences between the results of the present study and the findings of other investigators, additional points need to be considered.

For example, differences in training and experience of counselors could be an important factor. Two of the counselors in this study had had very little professional experience in the field of counseling. Another point is that the amount of information the students and counselors had available differed from other research projects. For instance, in the study by Young (7) the students had much more information to use in predicting their grades. A final point is the manner in which the students were asked to predict grades. In the study by Freehill (2), the students predicted grade point averages rather than indicating specific grades for a certain number of courses.

It should be mentioned that there are several reasons for the student and counselor estimates not having a higher predictive relationship to actual grades. One reason has to do with the aspiration level of students entering college. Torrance (6) has noted a strong trend for students to overestimate their abilities. A second reason is the observation by Budd (1) that counselors tend to err by overpredicting for the low achieving student. A third reason is the general unreliability of the grading system.

Nevertheless, inasmuch as multiple regression equations using high school grades and test scores have yielded fairly good predictions of college academic achievement, and since counselor estimates seem to show correlations with college grades comparable to the correlations of high school grades or of test scores with college grades, it might be worthwhile to add this variable of counselor estimates to the regression equations.

TABLE I

CORRELATION OF STUDENT SELF-ESTIMATES WITH COUNSELOR ESTIMATES AND OBTAINED GRADES

	df	chi square	correlation
Student pre-estimates vs. post-estimates	1	83.05	.55
Student pre-estimates vs. counselor estimates	6	30.53	.40
Student pre-estimates vs. actual grades	1	22.80	.38
Student post-estimates vs. counselor estimates	6	53.90	.50
Student post-estimates vs. actual grades	1	22.04	.37
Counselor estimates vs. actual grades	4	43.90	.50

Summary

The purpose of the study was to determine how well college freshmen and their counselors can predict first semester grades. Students predicted grades both before and after a precollege counseling and testing program. Counselor predictions were made at the end of each counseling interview.

A close correspondence was found between estimates made by students before and after the counseling program. While student post-estimates correlated higher with counselor estimates than did student pre-estimates, the correlation of both student estimates with actual grades was approximately the same. Counselor estimates correlated higher with obtained grades than did the student estimates. All correlations were statistically significant. The results were related to previous research concerned with the prediction of college achievement, using student and counselor estimates of grades.

REFERENCES

1. Budd, W. "Counselor versus Student Predictions of Grade Point Averages," Journal of Counseling Psychology, 1956, 3, 68-70.
2. Freehill, M. "Student Self-estimates as Guidance in Selecting Courses," College and University, 1952, 27, 233-242.
3. Matteson, R. "Self-estimates of College Freshmen," Personnel Guidance Journal, 1956, 34, 280-284.
4. Robertson, M. "Similarities and Differences Between Counselor and Student Reports of Counseling Interviews," Journal of Counseling Psychology, in press.
5. Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.
6. Torrance, E. "Some Practical Uses of a Knowledge of Self-concepts in Counseling and Guidance," Educational and Psychological Measurement. 1954, 14, 120-127.
7. Young F. "College Freshmen Judge Their Own Scholastic Promise, Personnel Guidance Journal, 1954, 32, 399-403.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

THE RELATIONSHIP OF CERTAIN FACTORS TO MARKING PRACTICES IN INDIVIDUALIZED REPORTING PROGRAMS

JOSEPH W. HALLIWELL
St. John's University
Jamaica, New York

Purpose

It was the purpose of this study to determine the the relationship of I. Q., academic achievement, and learning efficiency to report card grades in a district which had adopted an "individualized" type of marking program in the elementary schools. More specifically, this investigation sought to ascertain the relationships of I. Q., achievement as measured by a standardized test, and learning efficiency to course success or failure in the fourth and sixth grades.

Background

Within recent years there has been a marked tendency to substitute a system of evaluating pupil progress on the basis of the pupil's ability rather than the more conventional system of grading on the basis of class or national norms (7, 10). This type of marking procedure in which the individual is graded on the basis of the quality of his academic performance in relation to his academic aptitude rather than on the quality of his performance in relation to some group standard is often referred to as an "individualized" marking system. The advocates of the "individualized" program of reporting believe that such a reporting of pupil progress is more conducive to the social, emotional, and academic development of pupils (2).

Despite the fact that the "individualized" program is becoming increasingly popular and that certain prominent educators have maintained categorically that elementary school pupils should be graded on an individualized basis (3, 5, 10), few investigators have conducted any research on this newer type of reporting practice (7). A survey of the literature revealed that the only facet of the newer type reporting technique which had received adequate investigation was the relationship of "individualized" marking practices in the elementary school to subsequent academic success in the junior high school. The two studies conducted in this area sought to compare the effects that the "individualized" grading system and the traditional grading procedures in the elementary school had on junior high school achievement. The first investigation was concerned with achievement as measured by standardized tests (1), and the second was concerned with achievement as evaluated by teachers' marks (2). In neither study were any statistically significant differences found between the scores of the students who had been graded on an "individualized" basis and those who had been graded on a traditional basis in the elementary schools. One possible explanation of these findings is that neither of these methods of marking is significantly superior to the other in motivating students to achieve. A second possible explanation is that no significant differences in achievement were discovered because the marking practices of the two groups were not essentially different irrespective of the stated differences in marking philosophy. The second possibility can be explored by determining the relationship between intelligence, achievement, and learning efficiency and the satisfactory or unsatisfactory grades obtained by students in an "individualized" reporting program.

In an "individualized" program a student is not graded on the basis of some objective norm, but rather on the basis of the class work he produces in relation to his ability. In theory, such a method of evaluating pupil progress should not be related to I..Q. since we should expect as many dull pupils to actualize their potentialities as do bright pupils. So too, this technique for reporting pupil progress should not be related to achievement on standardized tests since there is such a high correlation between intelligence and achievement test scores (6). Theoretically, there should be a highly significant correlation between grades received in an "individualized" program and some measure of learning efficiency which appraises achievement in relation to intelligence. The present study seeks to determine whether grading in a district which has adopted an individualized marking program is most closely related to I. Q., achievement or learning efficiency. The finding of this investigation should demonstrate whether current marking practices in a district which has adopted a newer type of program for reporting pupil progress are significantly different from marking practices extant in the more traditional reporting programs.

Procedure

The investigation was conducted in a public school system in suburban Long Island. The district in which the study was undertaken had adopted the “individualized” system of reporting pupil progress in the elementary schools. The subjects of this study included all the fourth and sixth grade pupils in the district with the exception of those who were not present when an intelligence test was administered, those who were not present when an achievement battery was administered, and those who entered the term late. There were 665 students in twenty-nine classes in the fourth grade and 590 students in twenty-one classes in the sixth grade.

The Pintner Intermediate Test: Form A (Verbal) and the Stanford Achievement Test, Intermediate Battery - complete, Form D were administered to the subjects. A deviation I. Q. and a median grade equivalent were calculated for each student The report cards of the pupils were analyzed in order to determine the students who had received satisfactory marks and the students who had received unsatisfactory marks. Coefficients of correlation between I. Q. and median grade equivalents in achievement were calculated at both the fourth and sixth grade levels. Regression equations of achievement on intelligence were then set up for the fourth and sixth grades and an effort quotient was determined for each individual student. The effort quotient was obtained by dividing the actual achievement score by the predicted achievement score, and multiplying by 100 to remove decimals (12). The effort quotient was to serve as the measure of learning efficiency. The validity and the limitations of such a technique for measuring accomplishment in terms of ability have been the subject of recent research (4).

Separate bi-serial coefficients of correlation were calculated between intelligence and satisfactory or unsatisfactory grades, achievement and satisfactory or unsatisfactory grades, and learning efficiency and satisfactory or unsatisfactory grades at the fourth and sixth grade levels for boys and girls. A student was considered to have received unsatisfactory grades when he received an unsatisfactory grade in one or more academic areas. The determination of separate correlation coefficients for boys and girls was considered imperative because of the significant relationship between the sex factor and teacher grades (9). Bi-serial rather than the product moment coefficient of correlation was employed because the grading system is dichotomous. Satisfactory or unsatisfactory marks are the only possible grades.

The Findings

The magnitude of the bi-serial coefficients of correlation between I. Q. and satisfactory or unsatisfactory grades at the fourth grade level was .53 for the boys and .59 for the girls. At the sixth grade levels the correlations were .30 for the boys and .37 for the girls. All of these correlations were statistically significant at the .01 level of confidence.

The magnitude of the bi-serial coefficients of correlation between median grade achievement scores and satisfactory and unsatisfactory grades at the fourth grade level was .54 for the boys and .59 for the girls. At the sixth grade level the correlations were .35 for the boys and .41 for the girls. All of these correlations were of sufficient magnitude to be considered statistically significant at the .01 level of confidence.

The coefficients of correlation between learning efficiency and satisfactory and unsatisfactory grades at the fourth grade level were -.01 for the boys and +.02 for the girls. At the sixth grade level the correlations were -.01 and -.02. None of these correlations were found to be statistically significant.

Conclusions

Based on a consideration of the findings of this investigation certain conclusions seem warranted.

1. Intelligence is significantly related to grading practices at the .01 level of confidence in a district which has officially adopted an “individualized” grading program. The magnitude of the correlations between intelligence and grading in this study closely approximates the usual findings concerning the relationship between intelligence and school marks under a traditional grading system (8).
2. The correlations between median grade achievement scores and grading practices were slightly higher than those between intelligence and grades. This finding closely parallels the results of investigations employing traditional marking procedures (11).
3. There was no relationship between learning efficiency and grading practices.
4. The significant positive relationships between intelligence and grading, and standardized achievement scores and grading indicate that teachers employed in a school system which had adopted an “individualized” program of marking continued to evaluate students in a traditional manner irrespective of the stated marking philosophy of the district.
5. Any investigations which seek to determine the superior motivational effectiveness of different types of marking procedures must first determine whether the teachers involved in the research actually are adhering to the stated marking philosophies

of their districts.

REFERENCES

1. Baker, R. L., and Doyle, R. P., "A Change in Marking Procedure and Scholastic Achievement," Educational Administration and Supervision, April 1957, 43, 223-32.
2. Baker, R. L., and Doyle, R. P., "Elementary School Marking Practices and Subsequent High School Achievement," Educational Administration and Supervision, May 1958, 44, 158-166.
3. Caswell, H. L., and Foshay, A. W., Education in the Elementary School. New York: American Book Company, 1957, 292-93.
4. Coleman, W., and Ward, A. W., "Further Evidence of the Jangle Fallacy," Educational and Psychological Measurement, Winter 1956, 16, 524-26.
5. Elsbree, W. S., Pupil Progress in the Elementary School. New York: Bureau of Publications, Teachers College, Columbia University, 1943, 68-69
6. Kelley, L. L., Interpretation of Educational Measurements. Yonkers-on-Hudson, New York: World Book Company, 1927, 196-97.
7. Phillips, B. N., "Characteristics of Elementary Report Cards," Educational Administration and Supervision, November 1956, 42, 385-97.
8. Ross, C. C., and Stanley, J. C., Measurement in Today's Schools. New York: Prentice-Hall Inc., 1954, 115-16.
9. Schinnerer, M. C., "Failure Ratio: 2 Boys to 1 Girl," Clearing House, January 1944, 18, 263-65.
10. Thomas, R. M., Judging Student Progress. New York: Longmans, Green, 1954, 320-23.
11. Travers, R. M., Educational Measurement. New York: The Macmillan Company, 1955, 386-88.
12. Tsao, F., "Is AQ or F Score the Last word in Determining Individual Effort?" Journal of Educational Psychology, December 1943, 34, 513-25.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 2, October 1960)

SOCIAL STATUS AND ENROLLMENT IN TRADITIONAL MATHEMATICS

MARLOW A. MARKERT
The University of Tulsa

NUMEROUS RESEARCH investigations have evidenced that the socio-economic status of an adolescent's family is a factor significantly related to the type of curriculum which he pursues in school. These studies, however, have generally failed to consider the hypothesis that such a relationship may reflect primarily a common relationship to intelligence. Consequently, the relationship of socio-economic status and curriculum choice needs to be analyzed when the effects of intelligence are statistically controlled. This article reports the results of such an analysis. More specifically, the investigation tested the hypothesis that the number of high-school semester credits earned by a student in traditional mathematics is significantly related to the socio-economic status of his family, when measured intelligence is controlled.

Procedures

Every fifth member of the 1958 graduating class was selected from each of eight urban high schools located in Northeast Oklahoma. The resulting sample consisted of 226 males and 229 females. From official school records the following data were obtained for each of the selected graduates: (1) sex, (2) the specific traditional mathematics courses for which the student received high-school graduation credit, (3) parental occupation, and (4) the most recent intelligence quotient. The data were collected, tabulated, and analyzed separately for each sex.

The number of semester credits earned in traditional mathematics was totaled for each student, and constituted the criterion. Socio-economic status, the prediction variable, was measured by the prestige of parental occupation. According to the North-Hatt occupational prestige value index, each graduate was assigned a socio-economic status value. Measured intelligence constituted the control variable. To test the hypothesis, partial correlations between the number of mathematics credits and socio-economic status values, with intelligence held constant, were computed for each sex. The fiducial limits of the obtained correlations were established at the 5 percent level by the r to z transformation.

Results and Conclusions

When the intelligence factor was held constant, the partial correlations were .59 and .17 for males and females respectively.* The fiducial limits at the 5 percent level were $.50 < r < .67$ for boys and $.04 < r < .29$ for girls. Although the sex variable is probably the most significant factor associated with the girls' choice of enrollment in the mathematics curriculum, the fiducial limits of the obtained r values did not include zero for either male or female students. Therefore the conclusion was warranted that the number of high-school semester credits earned by a student in traditional mathematics is significantly related to the socio-economic status of his family, when measured intelligence is controlled.

The findings of this study indicate that the relationship of social status and curriculum choice does not reflect a common relationship to intelligence. The socio-economic status of a youth makes an independent contribution in the selection of the curriculum which the student pursues in school. If socio-economic status is also significantly related to enrollments in other elective subjects, when intelligence is controlled, the simple solution of providing multiple programs in the secondary school curriculum is extremely complex. So long as students of differing social status groups respond differently to the values of various school subjects, the designation of a specific program as appropriate for students of a specific intelligence level may express a lack of sophistication of social realities.

---

* The obtained r's between the number of mathematics credits and measured intelligence, with social status partialled out, were .57 and .33 for males and females respectively.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 NOVEMBER, 1960 NUMBER 3

CONTENTS

Evidence from School Records on the Effectiveness of Ability Grouping—John W. French 83

Further Appraisal of the Manner in Which the American Press Handles Material Relevant to Education—Thomas F. Fleming 92

Criteria for Industry Sponsored Educational Materials and Selected Sources for Chemistry—Keith I. Roper and Stanley B. Brown 99

A Longitudinal Study of the Relationship of High School Foreign Language and Mathematics Study to Freshman Grades—Joseph V. West and Benjamin Fruchter 105

Incidental Learning in Foreign Language Learning—Paul Pimsleur 111

A Note on the Validity of Borow's College Inventory of Academic Adjustment—W. James Popham and Mary R. Moore 115

Faculty and Student Perceptions of Cheating—Jack R. Frymier 118

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

EVIDENCE FROM SCHOOL RECORDS ON THE EFFECTIVENESS OF ABILITY GROUPING*

JOHN W. FRENCH
Educational Testing Service, Princeton, New Jersey

A CONTROVERSIAL topic in education or training is the advisability of dividing students into classes so as to form homogeneous ability groupings at two or more levels. Without such groupings, it is often argued that the slower students prevent the faster ones from moving ahead at their optimum rate, while the slower students are subjected to work that is beyond them and are likely to become discouraged. Teachers often complain that a heterogeneous class is more difficult to teach and to handle. On the other hand, many arguments are raised against ability grouping. It is sometimes felt that the slower students can benefit from the stimulation of the faster ones. In the public schools, the parents of students in slower classes complain that their children are receiving inferior teaching. It seems likely that instruction both in Navy and in civilian schools could benefit by more information on this point.

While there are a few recent studies in this area, much of the experimental work is more than 20 years old. This does not appear to mean that the problem was settled 20 years ago, because the results of studies carried out in the twenties and thirties were conflicting. Different experimental designs, different tests, different teaching methods, and different learning materials all contributed to a confused situation. The evidence from studies carried out in grade schools and high schools seems slightly to favor ability grouping, particularly where adaptations of teaching method were made. However, homogeneous grouping sometimes had a detrimental effect. Several summaries have been made of these studies (Cornell, 1936; Miller, 1930; Otto, 1950; Otto, 1954; Rock, 1929; Turney, 1931). A review of the literature by Ruth B. Ekstrom (1959) will be issued as an ONR Technical Report.

While a good many of the studies that have been carried out supply useful information, this seems to be a situation where large numbers of students receiving many different kinds of training will need to be studied before really definitive generalizations become possible. The need for data in large quantities derived from many different educational situations leads to the hope that use can be made of existing files of data. Otherwise it is necessary to go to the expense and inconvenience to schools of arranging a large number of different experimental setups. It is the object, then, of the present research to find out what can be learned about the effectiveness of ability grouping from data already available in school records.

Requirements of the Data

For the purposes of this study, little restriction was put on the content of the training. It was satisfactory to deal with situations involving academic or practical learning, from books or apparatus, in the Navy or in civilian schools. However, to answer the question being asked, it is essential that the records have the following characteristics: (a) They must involve sufficient numbers of cases to permit the use of large-sample statistics; (b) a test of ability in the kind of training under consideration must have been given in advance of training to all students being compared; (c) range in ability on this test must be reasonably wide; (d) some students must be grouped by ability and preferably (but not necessarily) some students should not be so grouped; and (e) the same form of an achievement test in the field of study must be given at the conclusion of the training to all groups of students being compared.

Navy Sources

A search was made for suitable data at the U. S. Naval Training Center, Bainbridge, and the U. S. Naval Station, Norfolk.

---

* This research was supported by the Office of Naval Research. For taking time to discuss training and school problems and for generously making available the data analyzed in this study, acknowledgements are extended to the Commanding Officer and to Lt. Commander C. M. Younger and Lt. M. C. Neal at U. S. Naval Training Center, Bainbridge, Maryland; to the Commanding Officer and Lt. Clinton at U. S. Naval Station, Norfolk, Virginia, and to Mr. Leo Weitz, Principal, James Monroe High School, Bronx, New York.

At Bainbridge, the greatest obstacle to finding appropriate data was the system the Navy uses, calling for the selection of only well-qualified students for every course. On the basis of the General Classification Test, a student is assigned to a particular Naval school according to his ability. Those with insufficient ability do not qualify for the course, while those with more ability than necessary are assigned to a more demanding program. As a result, the sample of trainees taking any one course is highly homogeneous in the abilities represented by the test. Two exceptions to this generalization were investigated. The first was the recruit training which occurred before classification by ability was undertaken. The other was a remedial reading program which handled students that varied from those who were very retarded in reading to those who fell only slightly below the minimum standard.

At Norfolk Naval Training Center it was thought that there would be more hope for finding heterogeneous groupings in some of the schools, particularly the Radioman's School, since at that center men from the fleet, not recently tested for ability, were received for training in courses requested by them.

Bainbridge Recruit Trainees. For each of 147 recruit companies entering Bainbridge during an 18-month period in 1956 and 1957, the company means and standard deviations on the General Classification Test and the means on the Recruit Final Achievement Test were copied from Navy records. While much of recruit training is physical or manual in nature, the RFAT was a pencil and paper examination of the amount of information the trainees had absorbed in such things as Navy regulations, ship and aircraft recognition, safety rules, etc.

The GCT and RFAT company averages were found to have a correlation of .26. A scatterplot was made of the data with GCT plotted on the ordinate and RFAT on the abscissa. Different symbols were used on the plot to represent the heterogeneity of each company as indicated by the size of its standard deviation on GCT, which varied from 7.4 to 11.6. While it would be better to compare some classes intentionally selected so as to be homogeneous with others selected so as to be heterogeneous, the amount of unintentional variation in homogeneity indicated here is probably enough to reveal any important relationships with teaching effectiveness. The line showing the regression of GCT on RFAT was drawn. Effectiveness of instruction was defined for purposes of this study by the vertical deviation from this regression line. No relationship appears between heterogeneity of ability and effectiveness of instruction as defined. In the recruit training situation, this negative result is readily understandable, since no systematic effort was made to use different training methods or schedules for men of different ability levels. It may well be that the motivation and morale in each company became far more important to average company achievement than differences in heterogeneity.

Bainbridge Remedial Reading Trainees. As they entered the Navy, all recruits were given an objective reading test which has been prepared in several parallel forms. Those cadets not coming up to the minimum standard were given special remedial work intended to make it possible for them to serve as effective seamen. After 3, 5, or 7 weeks, the trainee would take a second form of the reading test. Those who received a passing score would proceed to regular recruit training; those failing to receive such a score would continue the remedial course until they were able to pass. At times of heavy recruitment, from two to four reading classes would form simultaneously and ability grouping was practiced. At times of light recruitment, only one class would start training at one time, and all levels of reading trainees would study together in that class. As a result, the standard deviation of initial test scores for some classes was more than six times as large as that for other classes.

Plots of initial against final reading test scores were made for four classes finishing in three weeks, 31 classes finishing in five weeks, and 12 classes finishing in seven weeks. Again different symbols were used for plotting classes having different levels of heterogeneity on the initial reading test. No relationship between heterogeneity and the effectiveness of instruction was observed.

With these data it was possible to find out whether the able trainees do particularly well in high-ability classes and whether the less able trainees do well in classes with other men at their own level. To answer these questions the trainees in five-week classes were divided into four quartiles on the basis of initial test score. The classes were divided into three groups on the basis of mean score for the initial test. The 12 groups defined by these divisions have a very variable number of cases, because relatively few able trainees fall in the low-level classes, and relatively few low-ability trainees fall in the high-level classes. Figure 1 shows a plot of the 12 groups comparing their mean scores on the first and second reading tests. The number of cases in each group is indicated in the figure. There is some evidence that membership in a slow class tends to inhibit achievement for the low-ability readers, but because of the inconsistencies in the data, no attempt was made to demonstrate significance.

Norfolk Data

The Naval schools at Norfolk were visited for the purpose of finding groups of trainees that were more heterogeneous than those at Bainbridge. As was hoped, the trainees at the Radioman's School were found to vary quite widely in ability. However, it was learned that code instruction, the only part of the instruction which could be measured effectively, was carried out in such a way that able trainees would not be held back and slow trainees would not be overwhelmed. The trainees received signals for code practice through earphones at whichever of three speeds was most suitable for them.

At the Radioman's School and at other training schools at Norfolk, it was the practice to allow slower students to drop back so as to repeat part of the work in classes that started later than they did. As a result, all classes were affected by shifting membership, so that it was not possible to evaluate the effectiveness of instruction under a particular unchanging set of conditions. In Metalsmith's and in similar schools the classes were so small that instruction could be made so individual that each student was allowed to work and to learn at a speed suitable to him. In a sense, then, it was the effectiveness of the Navy school system in providing for instruction at a suitable level that made the training data unusable in this investigation.

Public High School Data

Extreme homogeneity within a course or continually changing membership of classes are situations that seemed less likely to occur in civilian schools than in Navy schools. To locate appropriate data in public high schools, the purposes of the project were discussed with Dr. J. Wayne Wrightstone, Director of Research for the New York City schools. He suggested the James Monroe High School as one where suitable records would probably be available. At this school, all of the essential requirements of the data were present. The Nelson-Denny Reading Test and I. Q. were given to all students; ability grouping was practiced where possible; and major courses ended with a final uniform achievement test for all classes. Pertinent data were collected for the classes of 1958 and 1959 for tenth-grade courses in Biology, World History, and Geometry.

Intercorrelations of class means were computed for the two aptitude test scores and the end-of-course achievement test. These intercorrelations, means, and standard deviations tend to be high. Reading scores and I. Q. correlate with each other from .87 to .98. The predictors correlate with achievement from .65 to .94. It is to be remembered that these are correlations of ability-grouped class means rather than individuals. The I. Q. test is the better predictor for Biology and Geometry. The reading test is the better predictor for World History.

As was done with the Navy data, plots were made showing the regression of class means of predictor tests on class means for achievement. Figure 2 shows the plot of I. Q. and the Geometry achievement test scores. The homogeneity of the classes is indicated, and classes taught by the same teacher are noted by code letters. Plots for Biology and World History are omitted to save space. They resemble Figure 2 in that homogeneity seems entirely unrelated to teaching effectiveness as judged by deviations of class means from the regression line. As in Figure 2, there is some evidence in the other two plots that some teachers were consistently successful and some were consistently unsuccessful whether their classes were homogeneous or not. Figure 2, however, differs from the other two plots in showing the most thorough ability grouping. While the lack of relationship between homogeneity and teaching effectiveness seemed understandable for recruit trainees, some relationship was expected here, because the teachers, by lip service at least, adjust the teaching schedule to the ability of the class. Since this is more possible with homogeneous classes, the effectiveness of teaching was expected to be better with such classes.

In addition to finding that homogeneity of classes was unrelated to teaching effectiveness, it was again desirable, as in the Navy data, to find out whether able students achieved more in high-ability classes and slow students were better off in low-ability classes. As was done for the remedial reading trainees at Bainbridge, the data for the individual high school students taking each course were divided into quartiles on the basis of the predictor test score. The classes were divided into those with relatively high means and those with relatively low means on the predictor test score. Figures 3, 4, and 5 show plots of the eight resulting means for each course. Again it should be remembered that the number of classes is small for upper quartile students in low-ability classes and for lower quartile students in high-ability classes. The number of cases in each group is indicated in the figure.

Evident in Figures 3, 4, and 5 is the tendency for both able and slow students to achieve more in high-ability classes than they do in low-ability classes. This confirms the uncertain appearance of this tendency in Figure 1. The statistical significance of these differences was tested by analysis of covariance (Gulliksen, 1950). The regression of predictor scores on achievement for students in high-ability classes was compared with that for students in low-ability classes. Only the middle two ability quartiles were used in this analysis, because for them there is an adequate number of cases in both high- and low-ability classes. Furthermore, curvilinearity of regression was unlikely in the middle two quartiles. This method of analysis of covariance indicates significance of differences between the intercepts of the regression lines assuming equal errors of estimate for the groups and equal slopes of the regression lines. The intercepts were found to be significantly different for Biology and World History. However, the significance is in some doubt, because errors of estimate were also found to be significantly different.

Discussion and Conclusions

When testing hypotheses in the behavioral sciences, the pitfalls in using available data rather than controlled experimentation can be treacherous. In this investigation it was desired to reach some con-

FIGURE 1.

PLOT OF MEANS ON INITIAL AND FINAL READING TESTS FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG BAINBRIDGE READING TRAINEES

(The number of cases for each point is indicated on the plot.)

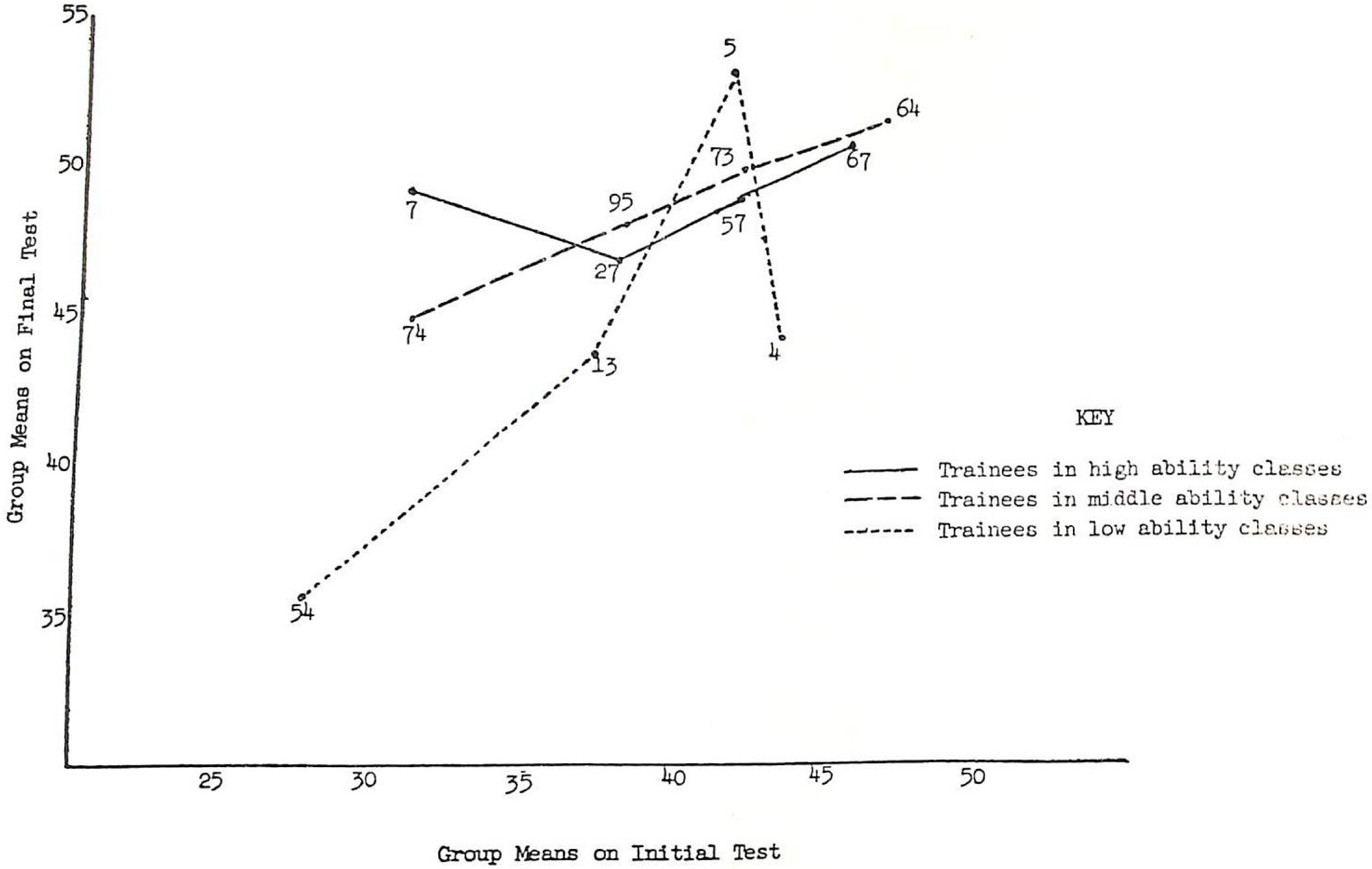
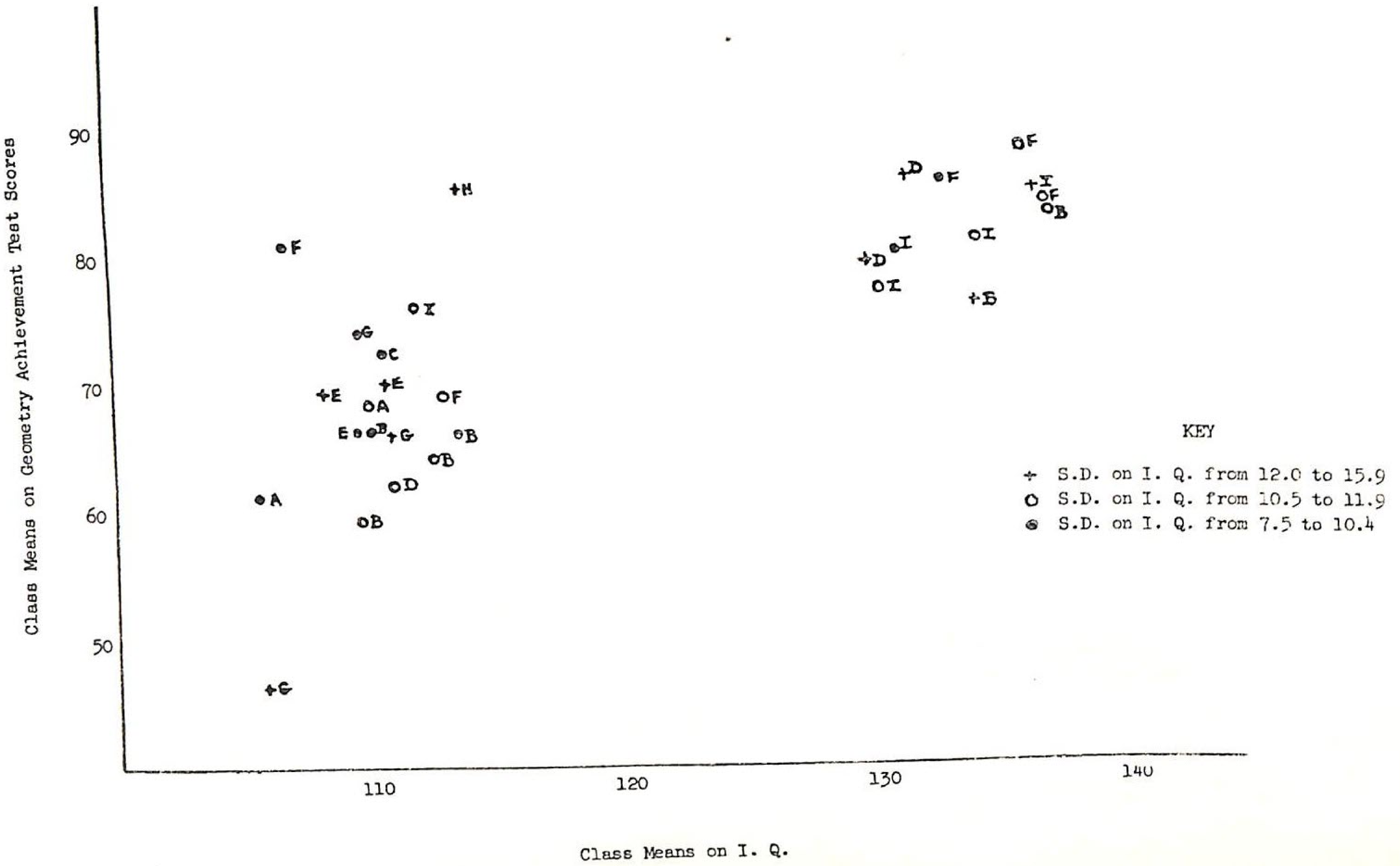


FIGURE 2.

PLOT OF HIGH SCHOOL CLASS MEANS ON I. Q.
AND GEOMETRY ACHIEVEMENT TEST SCORES



FRENCH

FIGURE 3.

PLOT OF MEANS ON I. Q. AND BIOLOGY ACHIEVEMENT TEST FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG HIGH SCHOOL STUDENTS

(The number of cases for each point is indicated on the plot.)

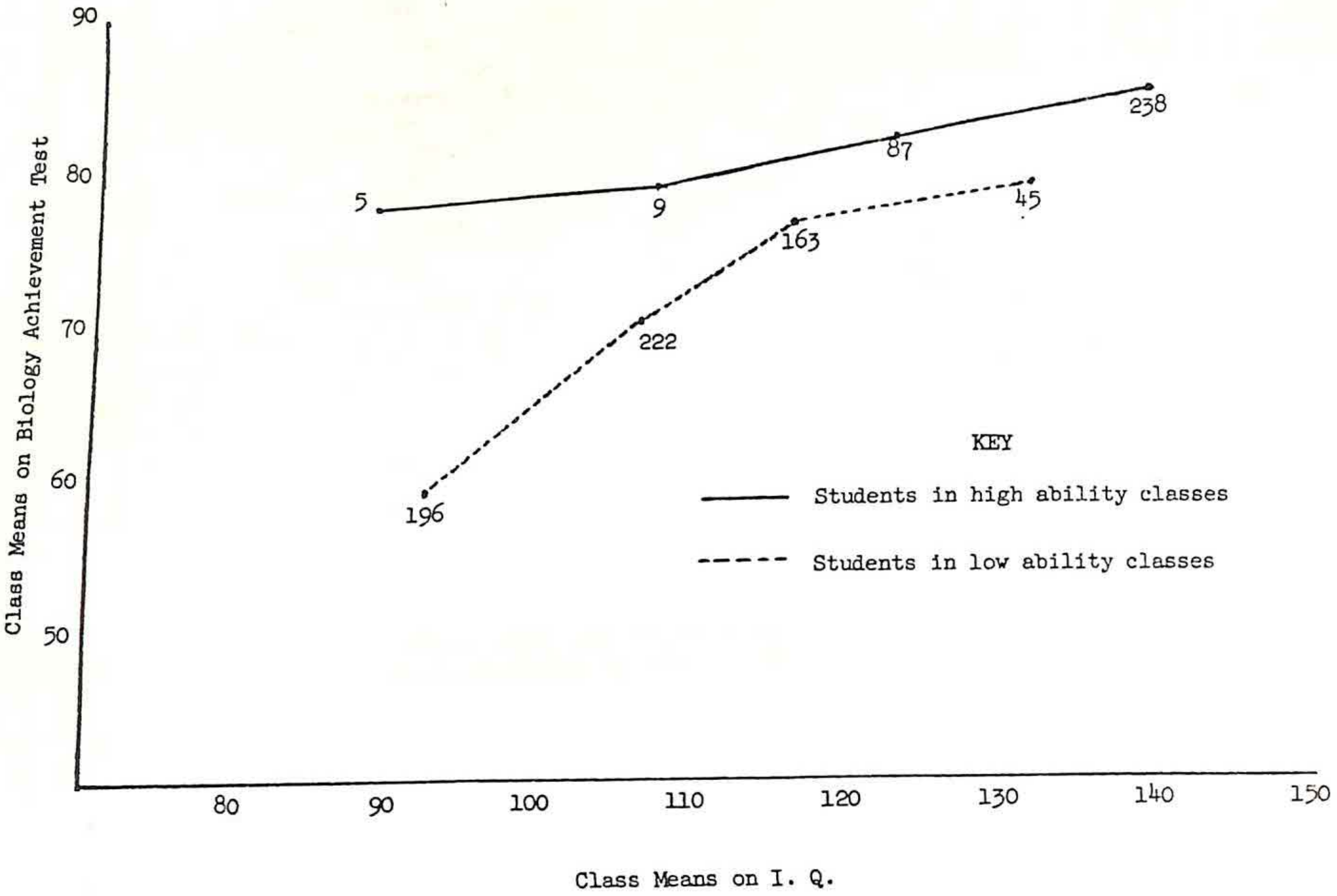
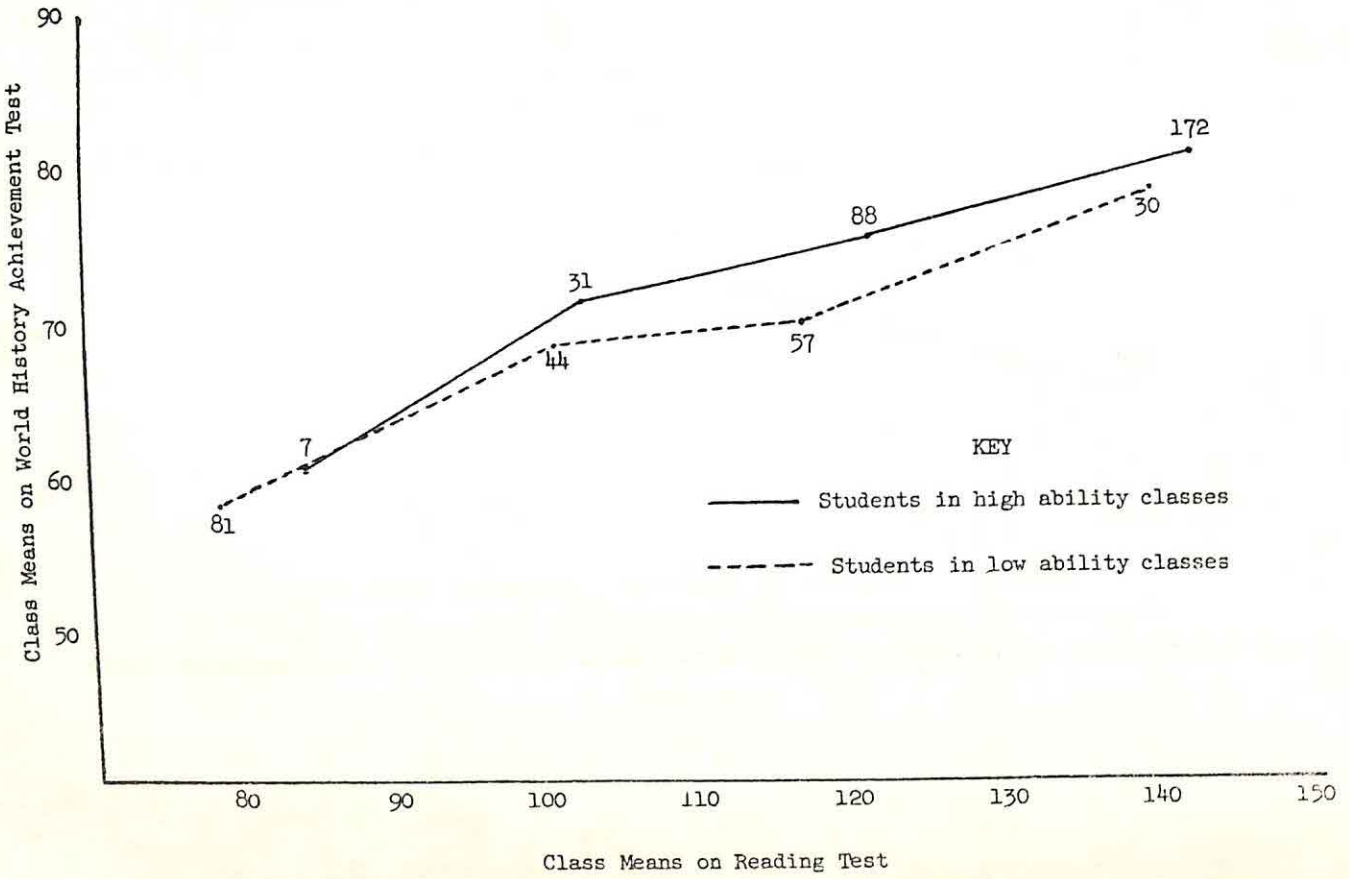


FIGURE 4.

PLOT OF MEANS ON READING TEST AND WORLD HISTORY ACHIEVEMENT TEST FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG HIGH SCHOOL STUDENTS

(The number of cases for each point is indicated on the plot.)



FRENCH

FIGURE 3.

PLOT OF MEANS ON I. Q. AND BIOLOGY ACHIEVEMENT TEST FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG HIGH SCHOOL STUDENTS

(The number of cases for each point is indicated on the plot.)

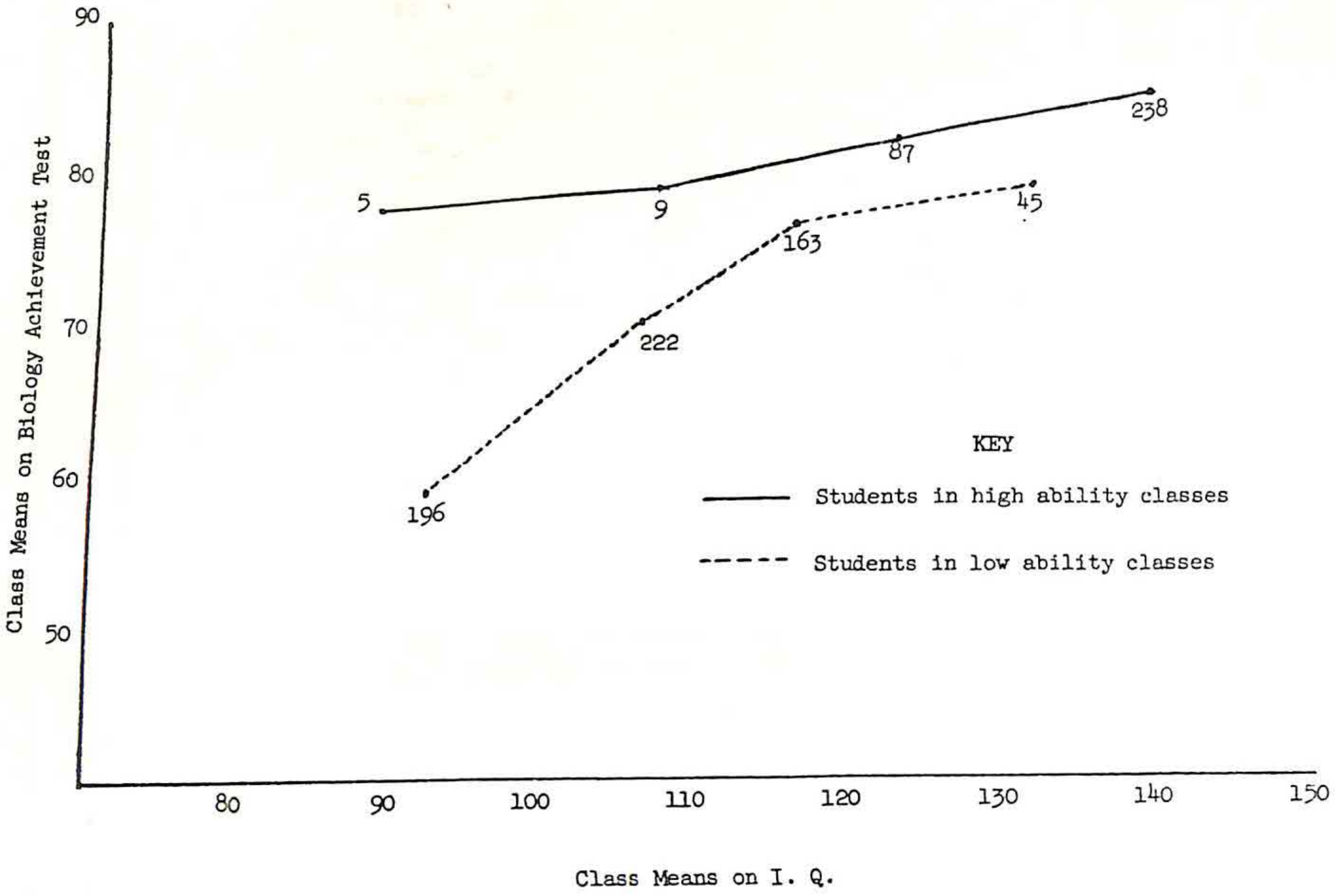


FIGURE 4.

PLOT OF MEANS ON READING TEST AND WORLD HISTORY ACHIEVEMENT TEST FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG HIGH SCHOOL STUDENTS

(The number of cases for each point is indicated on the plot.)

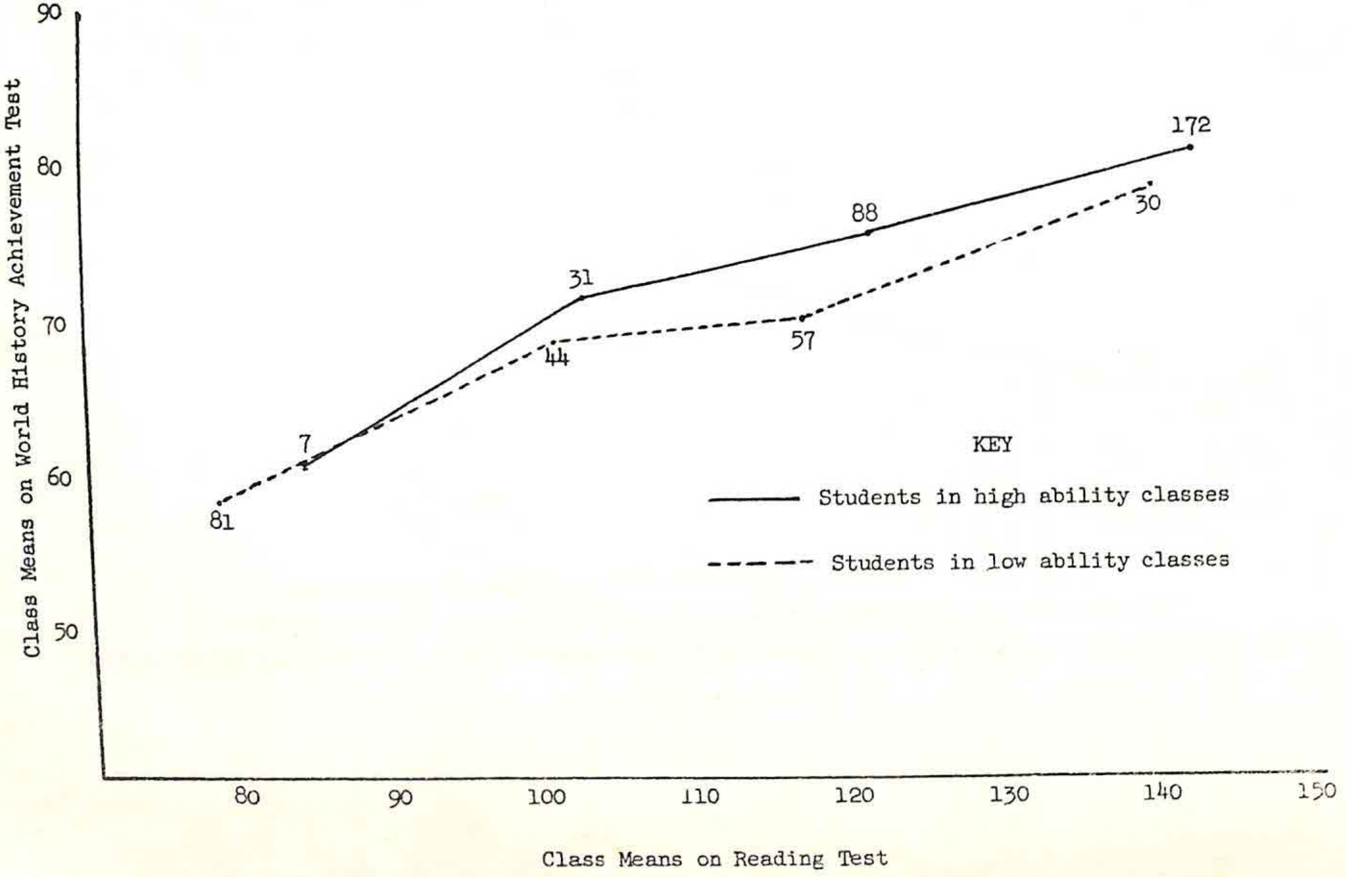
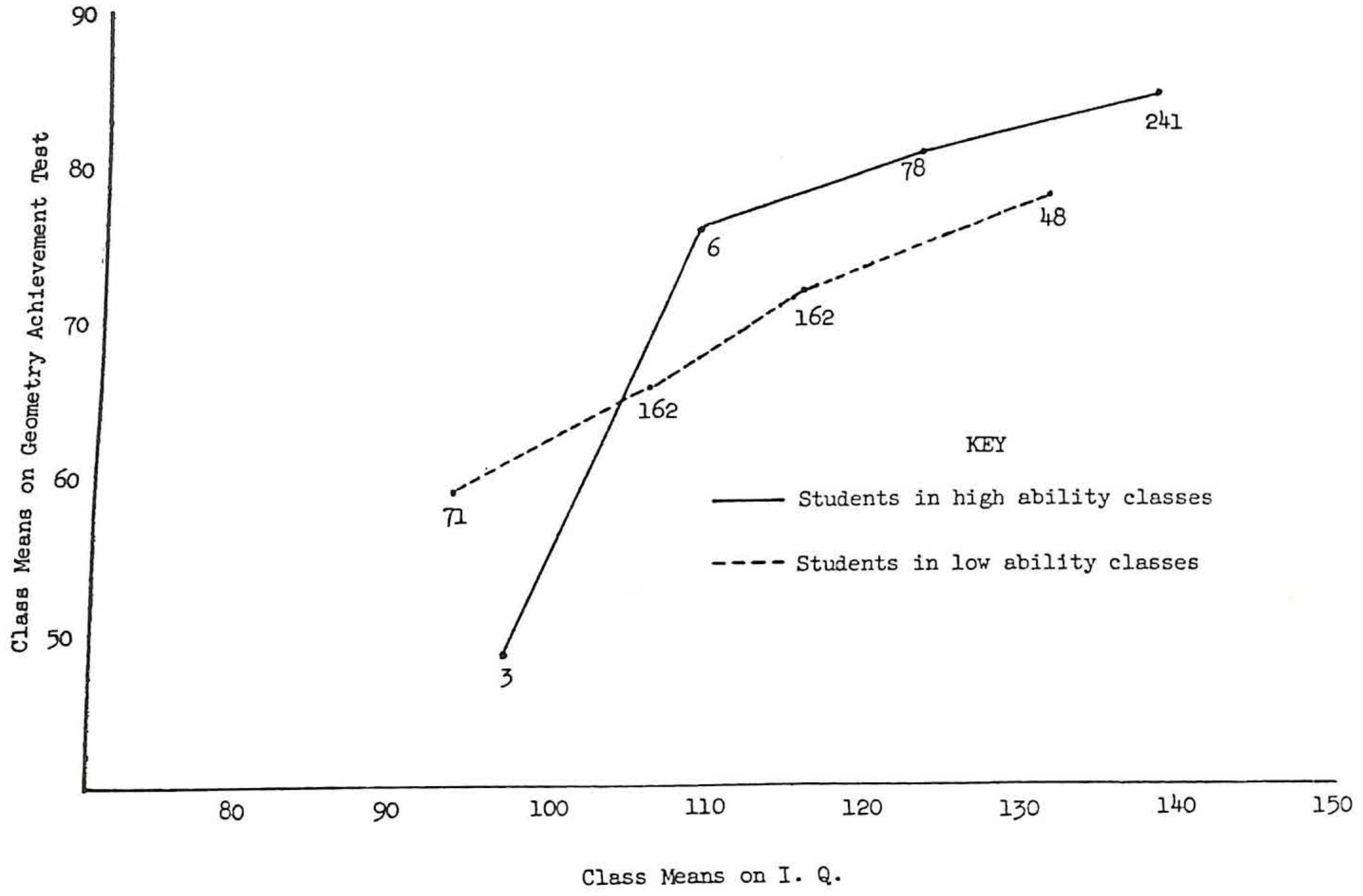


FIGURE 5.

PLOT OF MEANS ON I. Q. AND GEOMETRY ACHIEVEMENT TEST FOR APTITUDE QUARTILE AND CLASS ABILITY GROUPS AMONG HIGH SCHOOL STUDENTS

(The number of cases for each point is indicated on the plot.)



clusion on the effectiveness of teaching under varying conditions of homogeneity and ability grouping. In two situations in a Navy school the data seemed worth analyzing to obtain evidence on this problem. However, because of excess homogeneity within courses or because of continually changing membership of classes, it does not appear that many situations in the Navy are likely to be fruitful sources of the kind of data required. The availability of suitable data for this problem in public high schools seems more likely. However, it will be explained why findings based on high school records are very difficult to interpret. Nevertheless, data for three courses at one high school were analyzed.

In the analysis of both Navy and civilian data, there was found to be no relationship between achievement and homogeneity of ability within classes. Since ability grouping attempts to make classroom groups more uniform, this seems to be evidence that ability grouping does not by itself increase the effectiveness of instruction.

However, in the high school data there was clear evidence that bright children do better in high-ability classes and that slow children also do better in high-ability classes. This indicates that ability grouping is helpful to bright children, but harmful to most of the slow children, a finding that seems to confirm the feelings of parents. The parents of bright children are usually delighted to have their children in fast groups, while the parents of slow children often complain that ability grouping is unfair and undemocratic.

Unfortunately, an example of the pitfalls encountered in using available data may entirely account for the differences that were found. Ability grouping, as carried out at the high school participating in this study, is a complex process, as indeed it probably is at any school that attempts ability grouping. Previous grades, parents' and students' desires, scheduling difficulties, and other things play a part. It is likely that students who have been motivated enough to get high grades in previous years will be the ones chosen for the high-ability classes. Thus, it may be that achievement in the special classes is higher than would be predicted on the basis of aptitude, not because the instruction is better than in other classes, but simply because the students in those classes always have achieved better than predicted. Therefore, a definitive conclusion about ability grouping cannot be made on the basis of these data, although it may be fair to say that something has been gained in understanding of the way ability grouping operates in the schools.

In addition to setting up appropriate controls over the selection of students for experimental groups, the director of a definitive experiment would have to insure his readers that suitable teaching methods and schedules were used for high, low, and heterogeneous ability groups. Any differences found must be shown to have been made possible only by the ability grouping. There should be no possibility, for example, that inferior performance by heterogeneous groups could result from there being a better match between teaching methods and group needs for the homogeneous groups than there was for the heterogeneous groups.

The final conclusion must be that available data is no substitute for a controlled experiment. While a number of controlled experiments are reported in the literature, many of these are difficult to interpret for some reason. The problem seems to remain without an adequate answer.

Summary

An important question in education or training pertains to the advisability of dividing students into classes so as to form homogeneous ability groupings at two or more levels. This project represents an attempt to find evidence on this question in the files of Navy or civilian schools.

An absence of relationship was found between homogeneity of classes and the effectiveness of preparing students for achievement tests either in Navy schools or in a New York City high school. In the high school, however, students enrolled in high-ability classes appeared in some subjects to have a definite advantage over students in lower-ability classes, whether they were themselves fast or slow learners. However, since this finding may be accounted for by the subjective selection of students for the high-ability classes, it must be concluded that a definitive answer to the problems can only be obtained by means of controlled experiments.

REFERENCES

1. Cornell, E. L. "Effects of Ability Grouping Determinable from Published Studies," Thirty-Fifth Yearbook of the National Society for the Study of Education, 1936, 298-302.
2. Ekstrom, Ruth B. Experimental Studies of Ability Grouping; a Review of the Literature. Office of Naval Research Technical Report, 1959, in press.
3. Gulliksen, H. and Wilks, S. S. "Regression Tests for Several Samples," Psychometrika, 1950:15, 91-114.
4. Miller, W. S. and Otto, H. J. "Analysis of Experimental Studies in Homogeneous Grouping." Journal of Educational Research, 1930:21, 95-102.
5. Otto, H. J. Elementary School Organization and Administration. New York: Appleton-Century-Crofts, 1954, pp. 201-202.
6. Otto, H. J. "Homogeneous Grouping," in Monroe, W. (Editor), Encyclopedia of Educational Research. New York; MacMillan, 1950.
7. Rock, R. T., Jr. "A Critical Study of Current Practices in Ability Grouping," Catholic University of America, Educational Research Bulletin, 1929, Nos. 5 and 6.
8. Turney, A. H. "Status of Ability Grouping," Educational Administration and Supervision. 1931: 17, 21-42, 110-127.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

FURTHER APPRAISAL OF THE MANNER IN WHICH THE AMERICAN PRESS HANDLES MATERIAL RELEVANT TO EDUCATION

THOMAS F. FLEMING *
Stout State College
Menomonie, Wisconsin

IT WOULD seem essential, first, that one be aware of conflicts in opinion between educators and newspaper personnel as to what the public wants (or claims to want) to read about the educational process, in contrast to what the American press actually prints. Thus, a teacher reading his daily newspaper is likely to consider as overemphasized the coverage of school athletic events and personalities engaged in such activities, space given to scandal or controversy in school affairs, and publicity stories about administrators. What that teacher considers a lack of newspaper attention to the professional and social contributions of teachers and their profession often engenders within him a feeling of frustation and bitterness.

The following will be regarded by many teachers as being all too typical:

> One day last fall our student council sponsored "Operation Soap Suds." Nearly all the kids helped and we cleaned the school inside and out from top to bottom. The same day two of the older boys who played hookey stole a new Oldsmobile and were caught 20 miles out of town. The two got 18 inches of news space on the first page; the other 437 got one and one-half inches on page 36.[2]

The newspaper editor, on the other hand, is likely to hold different viewpoints. In regard to editorial support which newspapers over the years have given to teachers and educational progress, he tends to assume that the press has exerted a thoroughly positive influence. In fact, as far back as thirty years ago a national educational publication had already bestowed this accolade upon the American press for its services to the teaching profession:

> One doesn't go far in surveying the press before discovering that... teachers owe a large part of their increasing respectability to a generous recognition by the newspapers. 3

However, an editor's policy in printing news and pictures dealing with educational matters may have a distinctly different basis than that which influences the content of his editorial columns. To justify his use of certain material which he realizes will do harm to the educational process, he turns to any of a number of "definitions of news." Here are several widely-quoted examples:

> "News is whatever readers want to know about." 4
>
> "News is anything timely that interests a number of people." 5
>
> "It is the newspaper's duty to print the news--and raise hell." 6

Educators, in turn, react to what they consider poorly-balanced or unfavorable mass-media content. The resultant ill-feeling and breakdown in communication between educators and editors are typified in the following quotations which have appeared in national educational publications:

> ... public school teachers are condemned by the spokesman for the American public for failing to produce the kind of citizens this country needs and wants. 7

> The past two years (1949-50) have witnessed criticism in the press, periodicals, and other publications lavishing much attention upon groups and individuals who proclaim... children are being taught drivel. 9

---

* This appraisal results from the author's continuing interest in discovering what American newspapers report--in news stories, editorials and photos--about varied activities of pupils and teachers in our schools. His earlier investigations of American newspaper and magazine treatment of various aspects of teacher welfare were published previously in the Journal. (See footnote 1 at end of article.)

> The popular spirit of analyzing and criticizing teachers is beginning to be big business... it is the rare educator who does not feel self-conscious and wonder if he should see a psychoanalyst.[9]

Presentation of the data which follow is designed to go beyond isolated instances of supposed newspaper benevolence or malevolence in the handling of educational news and to describe, instead, the findings reported in certain relevant studies. One of the most frequently-quoted pioneers in the field of content analysis of mass-media material dealing with the teaching profession is Belmont M. Farley. His findings seem to indicate anomalies existent in the type of such material being presented to the reading public, rather than demonstration of bias. In "What to Tell the People About the Public Schools," Farley says:

> There is a tendency to supply readers of school news with the largest amount of information upon topics in which they have the least interest. The six topics of greatest interest to patrons get 25% of the news space while topics of lesser interest get 75% of news space.[10]

Farley's study included 5,000 persons in thirteen cities who indicated that the following six topics having to do with the teaching profession were of greatest interest to them: pupil progress and achievement, methods of instruction, pupil health, courses of study, value of education, discipline and behavior of pupils. Public interest in extra-curricular activities ranked at the very bottom of the long list of public educational interests. Yet, examination of 40,000 column inches of mass-media space devoted to those interests disclosed that 47.1% of that coverage dealt with extra-curricular interests, the educational phase of least public concern, whereas only 5.6% reported on the public's greatest interest in the teaching profession: the progress and achievement which pupils made under their teachers.[11]

On two later occasions, Farley is quoted in regard to his attitude on the manner in which the American press handles, in general, matters of an educational nature. In one instance he is reported as saying: "By and large the newspapers of the United States are friendly to the schools and that, as educational mediums themselves, they recognize the importance of freedom of expression, freedom to teach, and freedom to learn."[12]

In a personal letter to this writer, Farley examines, in more detail, both sides of the controversy:

> The teachers themselves have usually received fair and, in my opinion, adequate treatment by the press. By "fair" I mean they have taken into consideration all of the difficulties which handicap the effectiveness of teachers and have still given them credit for a highly useful service to society. By "adequate" I mean that they are not neglected. We have seen in the last twelve months many editorials pointing out the need for more teachers, better qualified teachers, and better paid teachers. We have seen a decline in the cartoon type of treatment in which teachers used to be quite frequently belittled and pictured as persons who might have stepped out of an album of 1835.[13]
>
> On the other side of the ledger, of course, we have to recall the too frequent accusations that teachers are left wing or subversive. Such charges do not come as a general rule from the press and do not represent editorial policy. Someone makes a wild statement about "subversive teachers" and gets quoted in the press.[14]

A widely-read study[15] in which W. Marvin Kemp examined editorial pages of daily and weekly newspapers in South Dakota indicates that South Dakota editors have made a clear-cut acceptance of their responsibilities as an extra-legal appraisal agency of the school. Kemp discovered that 35.03% of newspaper editorial material describing what teachers teach, how they teach it, and the results they obtain was written by editors. Yet, he indicated, only 5.3% of the items concerning the executive function of school organization--and these very frequently submitted to the papers by educators or publicity departments of school systems--could be classified as dealing with five of the six topics in which Farley found school patrons to be most interested. Table I offers a comparison of findings of other relevant studies.

Thus Kemp felt that mass-media editors, at least in their editorial columns, correlate material on education more closely with reader interest than do educational releases generally written by school personnel themselves. Kluckholn and Peterson substantiate responsibility for preparation of school news, claiming that 66 2/3% of such news is prepared within school systems.[16] A similar investigation by Rollo G. Reynolds indicates that in 83% of school systems sampled over the nation, superintendents furnished the major portion of that news.[17]

Benjamin Fine, nationally-syndicated writer on educational matters, is reported as having surveyed 180 college publicity officers to learn what these individuals, among whose main responsibilities was to provide news releases to editors, considered preferred items of an educational nature. Again, releases publicizing events of an extra-curricular nature were not considered as being worthy of top priority:

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

FURTHER APPRAISAL OF THE MANNER IN WHICH THE AMERICAN PRESS HANDLES MATERIAL RELEVANT TO EDUCATION

THOMAS F. FLEMING *
Stout State College
Menomonie, Wisconsin

IT WOULD seem essential, first, that one be aware of conflicts in opinion between educators and newspaper personnel as to what the public wants (or claims to want) to read about the educational process, in contrast to what the American press actually prints. Thus, a teacher reading his daily newspaper is likely to consider as overemphasized the coverage of school athletic events and personalities engaged in such activities, space given to scandal or controversy in school affairs, and publicity stories about administrators. What that teacher considers a lack of newspaper attention to the professional and social contributions of teachers and their profession often engenders within him a feeling of frustation and bitterness.

The following will be regarded by many teachers as being all too typical:

> One day last fall our student council sponsored "Operation Soap Suds." Nearly all the kids helped and we cleaned the school inside and out from top to bottom. The same day two of the older boys who played hookey stole a new Oldsmobile and were caught 20 miles out of town. The two got 18 inches of news space on the first page; the other 437 got one and one-half inches on page 36.[2]

The newspaper editor, on the other hand, is likely to hold different viewpoints. In regard to editorial support which newspapers over the years have given to teachers and educational progress, he tends to assume that the press has exerted a thoroughly positive influence. In fact, as far back as thirty years ago a national educational publication had already bestowed this accolade upon the American press for its services to the teaching profession:

> One doesn't go far in surveying the press before discovering that... teachers owe a large part of their increasing respectability to a generous recognition by the newspapers. [3]

However, an editor's policy in printing news and pictures dealing with educational matters may have a distinctly different basis than that which influences the content of his editorial columns. To justify his use of certain material which he realizes will do harm to the educational process, he turns to any of a number of "definitions of news." Here are several widely-quoted examples:

> "News is whatever readers want to know about." [4]
>
> "News is anything timely that interests a number of people." [5]
>
> "It is the newspaper's duty to print the news--and raise hell." [6]

Educators, in turn, react to what they consider poorly-balanced or unfavorable mass-media content. The resultant ill-feeling and breakdown in communication between educators and editors are typified in the following quotations which have appeared in national educational publications:

> ... public school teachers are condemned by the spokesman for the American public for failing to produce the kind of citizens this country needs and wants. [7]

> The past two years (1949-50) have witnessed criticism in the press, periodicals, and other publications lavishing much attention upon groups and individuals who proclaim... children are being taught drivel. [9]

---

* This appraisal results from the author's continuing interest in discovering what American newspapers report--in news stories, editorials and photos--about varied activities of pupils and teachers in our schools. His earlier investigations of American newspaper and magazine treatment of various aspects of teacher welfare were published previously in the Journal. (See footnote 1 at end of article.)

> The popular spirit of analyzing and criticizing teachers is beginning to be big business... it is the rare educator who does not feel self-conscious and wonder if he should see a psychoanalyst.[9]

Presentation of the data which follow is designed to go beyond isolated instances of supposed newspaper benevolence or malevolence in the handling of educational news and to describe, instead, the findings reported in certain relevant studies. One of the most frequently-quoted pioneers in the field of content analysis of mass-media material dealing with the teaching profession is Belmont M. Farley. His findings seem to indicate anomalies existent in the type of such material being presented to the reading public, rather than demonstration of bias. In "What to Tell the People About the Public Schools," Farley says:

> There is a tendency to supply readers of school news with the largest amount of information upon topics in which they have the least interest. The six topics of greatest interest to patrons get 25% of the news space while topics of lesser interest get 75% of news space.[10]

Farley's study included 5,000 persons in thirteen cities who indicated that the following six topics having to do with the teaching profession were of greatest interest to them: pupil progress and achievement, methods of instruction, pupil health, courses of study, value of education, discipline and behavior of pupils. Public interest in extra-curricular activities ranked at the very bottom of the long list of public educational interests. Yet, examination of 40,000 column inches of mass-media space devoted to those interests disclosed that 47.1% of that coverage dealt with extra-curricular interests, the educational phase of least public concern, whereas only 5.6% reported on the public's greatest interest in the teaching profession: the progress and achievement which pupils made under their teachers.[11]

On two later occasions, Farley is quoted in regard to his attitude on the manner in which the American press handles, in general, matters of an educational nature. In one instance he is reported as saying: "By and large the newspapers of the United States are friendly to the schools and that, as educational mediums themselves, they recognize the importance of freedom of expression, freedom to teach, and freedom to learn."[12]

In a personal letter to this writer, Farley examines, in more detail, both sides of the controversy:

> The teachers themselves have usually received fair and, in my opinion, adequate treatment by the press. By "fair" I mean they have taken into consideration all of the difficulties which handicap the effectiveness of teachers and have still given them credit for a highly useful service to society. By "adequate" I mean that they are not neglected. We have seen in the last twelve months many editorials pointing out the need for more teachers, better qualified teachers, and better paid teachers. We have seen a decline in the cartoon type of treatment in which teachers used to be quite frequently belittled and pictured as persons who might have stepped out of an album of 1835.[13]
>
> On the other side of the ledger, of course, we have to recall the too frequent accusations that teachers are left wing or subversive. Such charges do not come as a general rule from the press and do not represent editorial policy. Someone makes a wild statement about "subversive teachers" and gets quoted in the press.[14]

A widely-read study[15] in which W. Marvin Kemp examined editorial pages of daily and weekly newspapers in South Dakota indicates that South Dakota editors have made a clear-cut acceptance of their responsibilities as an extra-legal appraisal agency of the school. Kemp discovered that 35.03% of newspaper editorial material describing what teachers teach, how they teach it, and the results they obtain was written by editors. Yet, he indicated, only 5.3% of the items concerning the executive function of school organization--and these very frequently submitted to the papers by educators or publicity departments of school systems--could be classified as dealing with five of the six topics in which Farley found school patrons to be most interested. Table I offers a comparison of findings of other relevant studies.

Thus Kemp felt that mass-media editors, at least in their editorial columns, correlate material on education more closely with reader interest than do educational releases generally written by school personnel themselves. Kluckholn and Peterson substantiate responsibility for preparation of school news, claiming that 66 2/3% of such news is prepared within school systems.[16] A similar investigation by Rollo G. Reynolds indicates that in 83% of school systems sampled over the nation, superintendents furnished the major portion of that news.[17]

Benjamin Fine, nationally-syndicated writer on educational matters, is reported as having surveyed 180 college publicity officers to learn what these individuals, among whose main responsibilities was to provide news releases to editors, considered preferred items of an educational nature. Again, releases publicizing events of an extra-curricular nature were not considered as being worthy of top priority:

TABLE I

A COMPILATION OF CERTAIN STUDIES DESCRIBING THE PERCENTAGE OF NEWSPAPER EDUCATIONAL NEWS SPACE DEVOTED TO ATHLETICS AND EXTRA-CURRICULAR ACTIVITIES

Study	Athletics	Extra-Curricular
Hyde	54. 6	11. 4
Kemp	34. 7	33. 1
Grinnell	33. 3	No findings
Kluckholn	32. 0	21. 0
Reynolds	30. 0	No findings
Farley	27. 0	20. 1
Briscoe	25. 0	19. 0
Garlin	9. 9	11. 2

(Priority by rank)

1. Human Interest
2. Student Articles
3. Stories with Pictures
4. Extracurricular Activities
5. Science Stories
6. Educational Stories
7. National News
8. Surveys, Reports, Studies
9. Faculty Interviews
10. Scandal Stories [18]

Later, a subsidiary publication of the American College Public Relations Association repeated that poll among news and education editors of metropolitan newspapers. These newspaper people assigned even a lower priority to coverage of extra-curricular events:

(Priority by rank)

1. Human Interest
2. Stories with Pictures
3. Educational Stories
4. Surveys, Reports, Studies
5. Science Stories
6. Extracurricular Activities
7. Faculty Interviews
8. Student Articles
9. National News
10. Scandal Stories [19]

The aforementioned publication then polled a number of its own members to obtain their evaluation of the same categories of educational news. Taking into account the fact that the "Student Articles" category was inadvertently omitted from the questionnaire, the poll showed:

(Priority by rank)

1. Human Interest
2. Stories with Pictures
3. Extracurricular Activities
4. Science Stories
5. National News
6. Educational Stories
7. Scandal Stories
8. Surveys, Reports, Studies
9. Faculty Interviews [20]

However, in spite of the preceding evidence which seems to give strong indication that neither school publicists nor metropolitan newspaper editors are primarily interested in extra-curricular activities of schools, the data of two later studies agree with Farley's contention that a heavy ratio of American press coverage is devoted to those very activities. For instance, data contained in a Louisville University master's thesis, based upon study of a total of 89,984 lines of type reporting public school matters from July 1, 1948, to June 30, 1950, disclosed that 35% of those lines was devoted to some aspect of school athletics. [21]

The analysis by Robert C. Snider (22) involving all school photographs published during 10 months in 3,600 issues of 12 daily newspapers with a combined daily circulation of over 200,000, provides further disturbing data, especially if one accepts Snider's contention that photographs, particularly those of children, attract highest attention among newspaper readers. Snider discovered that the 12 newspapers had, in 10 months, published a total of 2,428 photographs of local public school activities. However, only 9% of all school photos discovered were of children under 12 years of age.

Variations were also evident in frequency-of-use of photos among the 12 newspapers examined. The two largest-circulating papers published more than half (1,252) of all photos used. Two others used a combined total of only 15 such pictures in ten months. One paper with a circulation of 6,704 printed 456 school photos in the ten-month period under examination; a paper with almost equal circulation (5,714) used only 52.

And what was depicted in the 2,428 photos analyzed? Why, 904 pictures (37%) were devoted to "athletics," whereas only .4% showed pupils in actual physical education classes from kindergarten through high school. The next largest category, 636 photos representing 26% of the total, was devoted to some phase of graduation. One is at first pleased that 133 photographs (6% of all photos) were of faculty members--until it is realized that 45% of these photos were of football or basketball coaches. One third of the papers used no faculty photos. The remaining 755 school photographs were scattered among 33 categories, 21 of which each received less than 1% of the total. "Photographs of pupils and teachers in classroom and laboratory situations," Snider reported, "are non-existent in most newspapers and quite rare in others."

Snider concluded his summation by urging improved press pictures of truly educational activities through cooperation between newspaper photographers and educators. The result would be decreased emphasis on photos having to do with athletics and fewer unwanted shots of "40 or 50 children row on row staring self-consciously from the page." Instead, there would become available reader-appeal photographs depicting what schools are actually doing.

However, the present situation, judged on the basis of the school photographs examined, Snider describes thusly:

> If our schools should be judged solely on the basis of the information conveyed by newspaper photographs of school activities, it would appear that the primary concern of these educational enterprises is a series of seasonal extracurricular activities crowded into a busy schedule of sporting events and climaxed with a vernal pageant called graduation. [22]

That Snider's reaction is in agreement with a criticism by D. G. Stout, writing in School and Community, can readily be noted:

> The fact is that the type of information constituting most of the school news leads easily to the assumption that school work consists of little besides athletic contests, dramatics, entertainment, forensics, and other extra-class activities. 23

And what data are available to indicate mass media's treatment of the individual teacher and of that teacher's problems within the teaching profession?

The "Study of School News in the Louisville Courier Journal from July 1, 1948, to June 30, 1950," to which earlier reference has been made in this analysis, found that 50% of news which dealt with educational matters during that two-year period was devoted to two topics: athletics (35%, as was mentioned before) and actions of the board of education (15%).

In descending order, the other 50% of total content discussed such topics as pupil activities of a non-athletic nature, materials concerning teachers specifically, and supervisory matters. 24 It was the contention of the author of the aforementioned study that, at least so far as her findings in certain issues of a single mass-media newspaper were concerned, the bulk of publicity given schools is favorable. However, she indicated that teachers, as such, are frequently criticized for their supposed failure in teaching the "3R's" and for teaching results in general. On the basis of her study, the author recommended that more mass-media coverage be devoted specifically to the accomplishments of teachers in the classroom.

A widely-discussed, documented investigation into criticisms which mass-media are actually printing about teachers and their activities was compiled in the early 1950's by the National Educational Association. 25 NEA's Research Division examined 256 items critical of educational endeavor which appeared in 168 different newspapers during the period August 22, 1951, through November 16, 1951. Newspapers of the District of Columbia and 46 states (not Missouri or South Dakota) were included. The 256 items of criticism comprised 60 critical news story items, 93 adverse editorial mentions, 47 detrimental references in signed columns and 56 disparaging allusions made by writers of "letters to the editor." Although the study embraced critical items directed at a number of phases of school operation, teachers in particular were singled out for heavy censure.

Data of the NEA study indicate that alleged subversive instruction by teachers--defined by the NEA as teaching about such matters as socialism, the New Deal, and the welfare state--elicited over one-half (53.3%) of unfavorable news comment discovered. Ranking second (30% of the news items) was critical news reference to teachers' success with the "3R's." As was the case with news items of a critical nature, NEA's compilation of newspaper editorials indicates that editorial reference to subversive instruction by teachers ranked first (44.1%) among categories surveyed. Teacher success with the 3 R's also received considerable editorial criticism (18.3%).

The NEA study also surveyed detrimental references appearing in certain articles by newspaper columnists, most such material of widely syndicated columnists being designated as "nationwide." Again, alleged subversive instruction by teachers was given the greatest attention (59.6%) by newspaper columnists. These columnists' critical allusions to teachers' methods (48.9%) and to teacher success with the 3 R's (46.8%) ranked second and third respectively.

Finally, the NEA survey examined the content of 56 "letters to the editor" appearing in 38 different newspaper issues. Writers of "letters to the editor" also were strongly motivated to allege that teachers were providing subversive instruction to American youth, for such allegations composed 21.4% of the letters surveyed--the same per cent devoted to lack of teacher success with the 3 R's. The only category of a higher total (25.0%) was that criticizing moral and spiritual values in America's public schools.

After warning that its survey is by no means to be interpreted without caution and reservation, NEA draws the following two conclusions:

1. that newspapers seem to be especially concerned about the possibility of some teachers spreading "subversive" propaganda, although the term "subversive" is never clearly defined.
2. that newspapers are concerned with the kind of education provided by this nation's teachers, teaching profession, and public schools.

This writer completed, at about the same time as the aforementioned NEA report, "A Study of Trends in Interest and Support of Teacher Welfare Exhibited by Prominent American Mass-Readership Publications at Three Significant Periods of American Life: 1931, 1941, 1951." 26 Using as a criterion of source selection those publications which are among America's largest-circulating newspapers and magazines, the study examined every news story and every editorial in over 9,000 daily newspaper issues and highest-circulating magazine issues of 1931, 1941, and 1951. As has been said earlier, results of that study were reported in detail in the following issues of the Journal: September, 1954; January, 1955; March, 1955.

In brief, findings from that research showed that newspapers and mass-circulation magazines of those eras were generally friendly (or in some cases indifferent) in news space and editorial space toward improved teacher salaries, guaranteed salary schedules for teachers, age of teacher retirement, pensions, sick benefits, vacations, and leaves. However, the content both of newspapers and of magazines examined revealed a trend toward increased criticism of teachers as being too frequently subversive and of being often incompetent in teaching the 3 R's.

Several later studies are also worthy of attention from persons interested in the public's image of the teacher. In 1957, NEA published a well-documented report, The Status of the American Public School Teacher, 27 which describes the male and female public school teacher as the public views such an individual and enumerates that teacher's typical professional, social, and civic contacts with the public.

A 1958 statement of the Problems and Policies Committee of the American Council of Education28 accepts the validity of "critical analysis of our educational system" but warns against public "name-calling and fault-finding." Particularly relevant to estimating the public's attitude toward the teacher and his profession--and the teacher's role in the formation of that attitude--is the 1959 NEA research monograph, The Classroom Teacher and Public Relations. 29 Of special significance to the analysis here under discussion are NEA's findings that although 48% of respondents to the questionnaire had within one year noted a classroom or school event or events sufficiently newsworthy to be called to the attention of school officials, only one teacher in four had described such an event in written form for use by mass media. Yet, as the report points out, "Many school activities make interesting reading... for the public." 30

Summary

This analysis described the author's continuing interest in the types of material which the American press prints about varied activities within the educational process and about the individual teacher and his or her achievements in that process. Much evidence was produced to indicate that the press is still emphasizing, especially in news stories and photos, the extra-curricular activities in schools and the personnel in such activities, in spite of the claim by readers and even metropolitan editors that such topics are not of primary concern. Nevertheless, the data of study after study reveal that varsity athletics, administrative actions or personnel, and similar non-classroom occurrences receive the greatest proportion of news and photo coverage. In-class news stories and features having to do with such supposedly-popular topics as pupil achievement and student human interest receive far less than their just due. That such a contradictory situation exists is even more illogical inasmuch as a number of studies quoted claim that much of the material--the wrong type of material--is prepared within school systems.

So far as treatment of teachers by the American press is concerned, evidence reported in this analysis indicates that this situation, too, is far from satisfactory. Too little favorable coverage seems to be afforded individual teachers and their accomplishments. Instead, news columns, editorial pages, syndicated columnists, and writers of "letters to the editor" warn that a significant number of teachers may be subversive, but often that term is never clearly defined or is used to "label" a teacher who merely devotes time in the classroom to the accomplishments of the now-past New Deal. Also, all the preceding departments of the press contain the rather constant complaint that teachers do not do a very satisfactory job of teaching the "3 R's."

The editors of a relatively-recent edition of The Yearbook of Education warn against the results of such continued unfavorable mass media material:

> We should ask ourselves.... whether the functions assigned to teachers can in fact be discharged by persons of low social status. Would a preacher despised by his hearers be accepted as a moral guide? Would a medical man respected no more than a domestic servant be trusted by his patients, and would he be able to obtain from the community the material resources needed to apply modern scientific knowledge? 31

To counteract the preceding danger, NEA's "The Classroom Teacher and Public Relations," after having noted that only one teacher in four annually prepares for mass-media an average of even one news story favorable to schools or teachers, offers perhaps the best suggestion--which may also be interpreted as a subtle rebuke:

> In order for this public relations resource to be tapped, the information must be written and distributed to the public through the local newspaper, radio, or in some other way. 32

FOOTNOTES

1. See three studies on mass-media treatment of various phases of teacher welfare, by Thomas F. Fleming, Journal of Educational Research: September, 1954; January, 1955; March, 1955.
2. National Education Association of the United States, Research Division, Teacher Opinion on Pupil Behavior, 1955-56 (Washington: National Education Association, April, 1956), pp. 59-60.
3. Nation's Schools, October, 1931, p. 19.
4. Stanley Johnson and Julian Harris, The Complete Reporter (New York: MacMillan Company, 1948), p. 19.
5. Definition often attributed to Willard G. Bleyer, late director of the School of Journalism of the

University of Wisconsin.

6. Wilbur F. Storey, editor of the Chicago Times (1861), quoted by L. R. Campbell and R. E. Wolseley in Newsman at Work (Cambridge: Houghton Mifflin, 1949), p. 23.
7. School and Society, July 14, 1951, p. 20.
8. School and Society, October 7, 1951, p. 263.
9. Nation's Schools, March, 1951, p. 39.
10. Belmont M. Farley, What to Tell People About the Public School (New York: Bureau of Publications, Teachers College, Columbia University).
11. Ibid.
12. Belmont M. Farley, quoted in the Bulletin of the National Association of Secondary School Principals, April, 1952, p. 339.
13. Belmont M. Farley, personal letter to the author, August 3, 1954.
14. Ibid.
15. W. Marvin Kemp, "A Weakness in School Publicity," Nation's Schools, October, 1941, p. 30.
16. Kluckholn and Peterson, ibid.
17. R. G. Reynolds, ibid.
18. As quoted in Vee-Eye Newsletter of the American College Public Relations Association, Don TourTelot, editor.
19. Ibid.
20. Ibid.
21. "A Study of School News in the Louisville Courier-Journal from July 1, 1948, to June 30, 1950." Unpublished master's thesis, Louisville University.
22. Robert C. Snider, "How Well Do We Picture the Schools?" Wisconsin Journal of Education, February, 1955, pp. 10-11.
23. D. G. Stout, Teacher and Community, quoted by Margery W. Steer, "The Citizens' Corner," The Journal of Education, October, 1951, p. 204.
24. "A Study of School News in the Louisville Courier-Journal from July 1, 1948, to June 30, 1950," op. cit.
25. National Education Association of the United States, Research Division, Current Criticisms of Schools and How They Are Being Met (Washington: National Education Association, February, 1952).
26. Thomas F. Fleming, "A Study of Trends in Interest and Support of Teacher Welfare Exhibited by Prominent Mass-Readership Publications at Three Significant Periods of American Life: 1931, 1941, 1951." Unpublished doctoral thesis, University of Wisconsin.
27. National Education Association of the United States, Research Division, The Status of the American Public-School Teacher (Washington: National Education Association, February, 1957).
28. Public Understanding and Support for Education, A Statement of the Problems and Policies Committee of the American Council on Education (Washington: Publications Division, American Council on Education, February, 1958).
29. National Education Association of the United States, Research Division, The Classroom Teacher and Public Relations, Research Monograph 1959-M2 (Washington: National Education Association, July, 1959).
30. Ibid., p. 19.
31. The Yearbook of Education (1953), "Status and Position of Teachers," (New York: World Book Company, 1953), pp. 7-8.
32. National Education Association of the United States, Research Division, The Classroom Teacher and Public Relations, op. cit., p. 19.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

CRITERIA FOR INDUSTRY SPONSORED EDUCATIONAL MATERIALS AND SELECTED SOURCES FOR CHEMISTRY

KEITH I. ROPER
The Martin Company
Denver, Colorado

STANLEY B. BROWN
University of California
Berkeley, California

FOR ALMOST a half-century American educators have been concerned with free and inexpensive instructional materials that have been made available by industry in increasing quantities. A recent review of the literature revealed opinions that too many business interests are competing with basic curricular essentials for the time and attention of teachers and students. Certain educators view these materials collectively and disapprove them as having little or nothing to offer except distorted or slanted information and advertising.

During the past fifty years or so, and more frequently in recent years, many educators and educational organizations have stated that many of the offerings of industry have educational value and many sponsored materials are exceptionally schoolworthy. Opinions have been expressed that frequently industry-sponsored instructional materials are in harmony with educational objectives and consonant with the very best in educational philosophy.

Industry, on the other hand, has not denied that it is motivated by self-interest in providing teachers with free or inexpensive instructional materials. However, this interest does not preclude the possibility that their materials may provide worthwhile educational experience.

The problem would not appear to lie in the quantity of material, but, in the development of a ready yardstick for its qualitative measurement. The increasing use and continuing concern regarding sponsored materials emphasizes a teacher obligation objectively and frankly to determine sponsored materials that can be effectively employed. In doing this the teacher can influence the nature of the materials that industry sponsors.

As part of a larger study on the utilization of industry-sponsored materials, a jury of eighteen competent full-time high school chemistry teachers was used to develop criteria for selecting the best of industry-sponsored materials in the field of chemistry.

The development of evaluating criteria required jury judgment on tentative criteria pertinent to the various aspects of the material. Tentative criteria that were developed included: authenticity and accuracy, educational appropriateness, technical quality, ready availability, desirability of "package units", and encouragement of students toward careers in science. To characterize each tentative criterion as specifically as possible, tentative subcriteria were required in some instances.

To quantify the judgments of the jury, a rating scale was devised for measuring the importance of possible criteria. The scale provided the following numerical values for categorizing tentative criteria:

5. Essential -- Highly important,
4. Desirable -- Worthy but not essential,
3. Unimportant -- Not important, neutral,
2. Questionable -- Not sure, of doubtful merit,
1. Undesirable -- Objectionable.

Derivation of a mean score for all jury responses on each criterion and subcriterion, combining judgments, required that each category of the rating scale be arbitrarily assigned a value range:

4.5 to 5.0 Essential,
3.5 to 4.4 Desirable,
2.5 to 3.4 Unimportant,
1.5 to 2.4 Questionable, and
0.0 to 1.4 Undesirable.

The investigators do not claim that all of the possible aspects of evaluating and selecting industry sponsored materials were included in the question-

University of Wisconsin.

6. Wilbur F. Storey, editor of the Chicago Times (1861), quoted by L. R. Campbell and R. E. Wolseley in Newsman at Work (Cambridge: Houghton Mifflin, 1949), p. 23.
7. School and Society, July 14, 1951, p. 20.
8. School and Society, October 7, 1951, p. 263.
9. Nation's Schools, March, 1951, p. 39.
10. Belmont M. Farley, What to Tell People About the Public School (New York: Bureau of Publications, Teachers College, Columbia University).
11. Ibid.
12. Belmont M. Farley, quoted in the Bulletin of the National Association of Secondary School Principals, April, 1952, p. 339.
13. Belmont M. Farley, personal letter to the author, August 3, 1954.
14. Ibid.
15. W. Marvin Kemp, "A Weakness in School Publicity," Nation's Schools, October, 1941, p. 30.
16. Kluckholn and Peterson, ibid.
17. R. G. Reynolds, ibid.
18. As quoted in Vee-Eye Newsletter of the American College Public Relations Association, Don TourTelot, editor.
19. Ibid.
20. Ibid.
21. "A Study of School News in the Louisville Courier-Journal from July 1, 1948, to June 30, 1950." Unpublished master's thesis, Louisville University.
22. Robert C. Snider, "How Well Do We Picture the Schools?" Wisconsin Journal of Education, February, 1955, pp. 10-11.
23. D. G. Stout, Teacher and Community, quoted by Margery W. Steer, "The Citizens' Corner," The Journal of Education, October, 1951, p. 204.
24. "A Study of School News in the Louisville Courier-Journal from July 1, 1948, to June 30, 1950," op. cit.
25. National Education Association of the United States, Research Division, Current Criticisms of Schools and How They Are Being Met (Washington: National Education Association, February, 1952).
26. Thomas F. Fleming, "A Study of Trends in Interest and Support of Teacher Welfare Exhibited by Prominent Mass-Readership Publications at Three Significant Periods of American Life: 1931, 1941, 1951." Unpublished doctoral thesis, University of Wisconsin.
27. National Education Association of the United States, Research Division, The Status of the American Public-School Teacher (Washington: National Education Association, February, 1957).
28. Public Understanding and Support for Education, A Statement of the Problems and Policies Committee of the American Council on Education (Washington: Publications Division, American Council on Education, February, 1958).
29. National Education Association of the United States, Research Division, The Classroom Teacher and Public Relations, Research Monograph 1959-M2 (Washington: National Education Association, July, 1959).
30. Ibid., p. 19.
31. The Yearbook of Education (1953), "Status and Position of Teachers," (New York: World Book Company, 1953), pp. 7-8.
32. National Education Association of the United States, Research Division, The Classroom Teacher and Public Relations, op. cit., p. 19.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

CRITERIA FOR INDUSTRY SPONSORED EDUCATIONAL MATERIALS AND SELECTED SOURCES FOR CHEMISTRY

KEITH I. ROPER
The Martin Company
Denver, Colorado

STANLEY B. BROWN
University of California
Berkeley, California

FOR ALMOST a half-century American educators have been concerned with free and inexpensive instructional materials that have been made available by industry in increasing quantities. A recent review of the literature revealed opinions that too many business interests are competing with basic curricular essentials for the time and attention of teachers and students. Certain educators view these materials collectively and disapprove them as having little or nothing to offer except distorted or slanted information and advertising.

During the past fifty years or so, and more frequently in recent years, many educators and educational organizations have stated that many of the offerings of industry have educational value and many sponsored materials are exceptionally schoolworthy. Opinions have been expressed that frequently industry-sponsored instructional materials are in harmony with educational objectives and consonant with the very best in educational philosophy.

Industry, on the other hand, has not denied that it is motivated by self-interest in providing teachers with free or inexpensive instructional materials. However, this interest does not preclude the possibility that their materials may provide worthwhile educational experience.

The problem would not appear to lie in the quantity of material, but, in the development of a ready yardstick for its qualitative measurement. The increasing use and continuing concern regarding sponsored materials emphasizes a teacher obligation objectively and frankly to determine sponsored materials that can be effectively employed. In doing this the teacher can influence the nature of the materials that industry sponsors.

As part of a larger study on the utilization of industry-sponsored materials, a jury of eighteen competent full-time high school chemistry teachers was used to develop criteria for selecting the best of industry-sponsored materials in the field of chemistry.

The development of evaluating criteria required jury judgment on tentative criteria pertinent to the various aspects of the material. Tentative criteria that were developed included: authenticity and accuracy, educational appropriateness, technical quality, ready availability, desirability of "package units", and encouragement of students toward careers in science. To characterize each tentative criterion as specifically as possible, tentative subcriteria were required in some instances.

To quantify the judgments of the jury, a rating scale was devised for measuring the importance of possible criteria. The scale provided the following numerical values for categorizing tentative criteria:

5. Essential -- Highly important,
4. Desirable -- Worthy but not essential,
3. Unimportant -- Not important, neutral,
2. Questionable -- Not sure, of doubtful merit,
1. Undesirable -- Objectionable.

Derivation of a mean score for all jury responses on each criterion and subcriterion, combining judgments, required that each category of the rating scale be arbitrarily assigned a value range:

4.5 to 5.0 Essential,
3.5 to 4.4 Desirable,
2.5 to 3.4 Unimportant,
1.5 to 2.4 Questionable, and
0.0 to 1.4 Undesirable.

The investigators do not claim that all of the possible aspects of evaluating and selecting industry sponsored materials were included in the questionnaire that was submitted to the eighteen jury members; however, they do believe that it included all of the significant aspects involved in the process.

A questionnaire and a stamped envelope with the investigator's address was given to each member of the jury. At this time the purpose of the questionnaire was explained to the jury members and their cooperation solicited. All of the questionnaires were returned. No follow-up letters were necessary.

Although the questionnaire was carefully constructed and refined, the question of reliability was considered. Would the responses of a second administration of the questionnaire lead to scores that, with the exception of a chance variation, remain the same? The investigators thought it advisable to make a statistical check on the reliability of ratings.

After considering several methods of testing reliability, the retest method seemed to be the most suitable. This was done by randomly selecting nine members of the original jury and asking them to respond a second time to the questionnaire. The nine members were selected by taking each second person on the list of jury members.

Again all teachers who were asked to respond did so. The assigned value for each category was used to derive a rating for each tentative criterion and subcriterion. Instead of using the total scores for each respondent on the questionnaire, it was believed that a more accurate test of reliability would result from totaling the scores of all respondents for each tentative criterion and subcriterion in the questionnaire, for each of the two trials, and then compute the correlation of the item scores. In the item analysis as described in Table I, total responses for each item by the nine respondents to the first and second tests are recorded and compared. Variations are noted as numerical differences from the means of total responses to all items. The relationship of the two responses was then tested by the use of product moment correlation, which resulted in a calculated coefficient of correlation of 0.89.

In testing the reliability of the obtained coefficient of correlation, the following limitations were recognized:[1*] (1) the obtained coefficient of correlation was not a true coefficient of the population, but rather an estimated one, and (2) the number of cases in this study was too small to expect a normal distribution.

R. A. Fisher's z-function formula was employed to test the reliability of the obtained coefficient of correlation. This formula was used after reading McNemar's discussion of the r to z transformation, which stated two distinct advantages as follows:

> (1) the distribution of z for successive samples is independent of the universe value, r; i.e., for a given N the sampling distribution will have the same dispersion for all values of r;
>
> (2) the distribution of z for successive samples is so nearly normal that it can be totaled as such with very little loss of accuracy.[2]

Garrett also discussed the z-function as a test of the reliability of the coefficient of correlation, a part of which read:

> A mathematically more defensible method of testing the significance of M, especially when the coefficient is high, is to convert the M into R. A. Fisher's z function and find the SE of r: (1) its sampling distribution is approximately normal and (2) its SE depends only on the size of the sample N, and is independent on the size of r.[3]

Using the formula

$$\sigma z = \frac{1}{\sqrt{16-3}} = \frac{1}{3.61} = 0.28$$

$$z = 1.47 \text{ (Table C in McNemar)}^{4}$$

$$\text{The } 0.95 \text{ confidence interval} = 1.47 \pm (1.96)(0.28) = 0.92 \text{ to } 2.02$$

Converted back to r's the confidence-interval was 0.73 to 0.97.

Thus the fiduciary probability was 0.95 that the true coefficient of correlation lies within the interval of 0.73 and 0.97 and may be expected in ninety five out of every one hundred times that the instrument is re-administered. Therefore it is believed that the instrument is reliable enough to provide a basis for evaluating and selecting industry-sponsored instructional materials.

However, it was important that some indication be given relative to the stability of the various parts of the questionnaire. Item Number 7 was arbitrarily selected and subjected to the same analytical procedure. The calculations of the coefficient of correlation between the raw scores on the first and the second trials are given in Table II. The r to z transformation was again used. The resulting fiduciary probability was 0.95 that the true coefficient of correlation lies between 0.09 and 0.96 and may be expected in ninety five out of every one hundred times that the instrument is re-administered.

Again it is noted that a correlation based on so few cases is likely to be very unstable.

The mean score and the number of responses made on the first test by all members of the jury for each tentative criterion and subcriterion rated essential or desirable may be found in Table III. The dispersion of responses is tabulated under the various categories.

Industrial sources for instructional materials applicable to high school chemistry are almost inexhaustible. Some realization of the quantity and variety of materials can be obtained by reviewing list-

---

* All footnotes will be found at the end of the article.

TABLE I

PRODUCT-MOMENT CORRELATION OF RAW SCORE TOTALS ON EACH ITEM FOR RESPONSES BY THE RE-TEST GROUP

Item No.	X	Y	x	y	x^2	y^2	xy
1	41	42	1.8	3.1	3.24	9.61	5.58
1a	41	40	1.8	1.1	3.24	1.21	1.98
1b	36	36	-3.2	-2.9	10.24	8.41	9.28
1c	37	35	-2.2	-3.9	4.84	15.21	8.58
1d	41	39	1.8	0.1	3.24	0.01	0.18
1e	39	40	-0.2	1.1	0.04	1.21	- 0.22
1f	41	42	1.8	3.1	3.24	9.61	5.58
2	45	45	5.8	6.1	33.64	37.21	35.38
2a	40	40	0.8	1.1	0.64	1.21	0.88
2b	42	39	2.8	0.1	7.84	0.01	0.28
2c	39	38	-0.2	-0.9	0.04	0.81	0.18
3	38	40	-1.2	1.1	1.44	1.21	- 1.32
4	35	34	-4.2	-4.9	17.64	24.01	20.58
5	39	40	-0.2	1.1	0.04	1.21	- 0.22
6	37	39	-2.2	0.1	4.84	0.01	- 0.22
7	36	33	-3.2	-5.9	10.24	34.81	18.88

16 = 627 = 622

= 104.44 = 145.76 = 109.34

M = 39.2 M = 38.9

$$r = \frac{xy}{\sqrt{x^2 y^2}}$$

$$= \frac{109.34}{\sqrt{104.44 \times 145.76}}$$

$$= 0.89$$

TABLE II

PRODUCT-MOMENT CORRELATION OF RAW SCORES ON ITEM NUMBER 7 FOR RESPONSES BY THE RE-TEST GROUP

	X	Y	x	y	x^2	y^2	xy
	4	3	0	-0.56	0	0.31	0.00
	3	3	-1	-0.56	1	0.31	0.56
	3	1	-1	-2.56	1	6.55	2.56
	5	5	1	1.44	1	2.07	1.44
	5	5	1	1.44	1	2.07	1.44
	4	4	0	0.44	0	0.19	0.00
	4	4	0	0.44	0	0.19	0.00
	3	4	-1	0.44	1	0.19	-0.44
	5	4	1	0.44	1	0.19	0.44
=	36	33			6.00	12.07	6.00
M =	4.0	3.56					

$$r = \frac{xy}{\sqrt{x^2 y^2}}$$

$$= \frac{6.00}{\sqrt{6\ (12.07)}}$$

$$= 0.71$$

$$\sigma z = \frac{1}{\sqrt{9 - 3}}$$

$$= 0.41$$

$$z = 0.89$$

The 0.95 confidence interval = 0.89 ± (1.96) (0.41)
= 0.09 to 1.69
Converted back to r's the confidence interval is
0.09 to 0.96

TABLE III

CRITERIA FOR INDUSTRY-SPONSORED EDUCATIONAL MATERIALS

Column Ref: 5 - Essential 4 - Desirable 3 - Unimportant 2 - Questionable 1 - Undesirable

Criteria and Subcriteria	Mean Score	No. responding in category				
		5	4	3	2	1
1 . The free paper and exhibit instructional materials that are provided to high school chemistry teachers by industry should be educationally appropriate.	4.61	11	7			
1a. Constructive -- These materials should contribute to the objectives of science education and promote the educational philosophy of the school.	4.56	10	8			
1b. Adaptable -- These materials should readily fit into the curriculum and adapt to different teaching methods.	4.39	5	11	2		
1c. Time -- These materials should relate recent developments in the curriculum.	4.06	5	9	4		
1d. Communicative -- These materials should present information in a way that is consonant with the vocabulary and understanding of high school students.	4.39	7	11			
1e. Specific -- These materials should limit the number of concepts or ideas which they contain. They should not be cluttered with too many objectives.	4.39	9	7	2		
1f. Concise -- These materials should present information effectively and efficiently.	4.67	12	6			
2 . The free paper and exhibit instructional materials that are provided to high school chemistry teachers by industry should be authentic and accurate.	5.00	18				
2a. Ethical -- These materials should be in harmony with American social and moral ideals.	4.67	14	3		1	
2b. Objective -- These materials should present information, facts, and opinions without bias.	4.62	14	3	1		
2c. Commercialism -- These materials should not be self-interested to the point that they become objectionable; however, the sponsor should be identified so that teachers and students will know the source of each item.	4.28	6	11	1		
3 . The free paper and exhibit instructional materials provided to high school chemistry teachers by industry should be the best available source of information. If a source other than industry treats the same information to better advantage then that source should have priority over the free material provided by industry. The fact that instructional material is free is not a valid reason for using it.	4.44	11	4	3		
4 . The free paper and exhibit instructional materials provided to high school chemistry teachers are more desirable if provided as a package unit, complete with teacher's guide.	3.94	3	12	2	1	
5 . The free paper and exhibit instructional materials provided to high school chemistry teachers by industry should maintain high standards of technical quality. The utilization of art, color, and other technical aspects should be imaginative and effective.	4.22	6	10	2		
6 . The free paper and exhibit instructional materials provided to high school chemistry should be readily available. Provision should be made for expendable items such as experiment and exhibit materials.	4.17	5	11	2		
7 . The free paper and exhibit materials that are provided by industry to high school chemistry teachers should encourage students toward a consideration of science teaching as a career.	3.94	5	8	4	1	

ings that occur in professional journals and periodicals, resource units for courses of study, textbooks, and catalogues of sources published by educational institutions and private enterprise.

From such listings two hundred seventy-five different sources were identified as providing or likely to provide materials for chemistry instruction. One hundred seventy-eight of these sources responded to a letter request by sending samples of the materials that they provide.

The authors conducted the study while at the University of Colorado, Boulder, Colorado, and do not claim to have tapped all sources of materials that are available; but are of the opinion that the significant sources were contacted and the materials they provide obtained.

Of the one hundred seventy-eight sources that develop materials, only fifty seven were selected by a jury on the basis of the above criteria. Of the fifty seven, twenty-three industrial sources were selected by five or more jury members. These are shown in the following list. Complete source addresses are given for the convenience of those desiring instructional materials that were selected as the best sponsored offerings in the field of chemistry.

1. Allied Chemical and Dye Corporation, 61 Broadway, New York 6, New York
2. Aluminum Company of America, Educational Department, 1256 Alcoa Building, Pittsburgh, Pennsylvania
3. American Iron and Steel Institute, 350 Fifth Avenue, New York 1, New York
4. American Petroleum Institute, 50 West 50th Street, New York 20, New York
5. Behr-Manning, Division of Norton Company, Troy, New York
6. Bituminous Coal Institute, Educational Department, Southern Building, Washington 5, D.C.
7. Corn Industries Research Foundation, Incorporated, 3 East 45th Street, New York 17, New York
8. E. I. DuPont de Nemours Company, Public Relations Department, Wilmington 98, Delaware
9. Thomas A. Edison, Incorporated, Edison Storage Battery Division, West Orange, New Jersey
10. Esso Standard Oil Company, 15 West 51st Street, New York 19, New York
11. General Electric Company, 1 River Road, Schenectady 5, New York
12. General Motors, Department of Public Relations, P. O. Box 177, North End Station, Detroit 2, Michigan
13. Illinois Water Treatment Company, Rockford, Illinois
14. Mallinckrodt Chemical Works, Mallinckrodt Street, St. Louis 7, Missouri
15. Manufacturing Chemists Association, Incorporated, 1625 Eye Street, N. W., Washington 6, D. C.
16. Monsanto Chemical Company, 1700 South Second, St. Louis 4, Missouri
17. Reynolds Metals Company, 2500 South 3rd Street, Louisville 1, Kentucky
18. W. A. Schaeffer Pen Company, 311 Avenue H. Ft. Madison, Iowa
19. Standard Oil Company, 909 17th Street, Denver 1, Colorado
20. Texas Gilf Sulpher Company, Incorporated, 75 East 45th Street, New York 17, New York
21. Union Carbide and Carbon Corporation, 30 East 42nd Street, New York 17, New York
22. U. S. Steel Corporation, 71 Broadway, New York 6, New York
23. Upjohn Company, Kalamazoo 99, Michigan

FOOTNOTES

1. Henry E. Garrett, Statistics in Psychology and Education (New York: Longmans, Green and Co., 1953), p. 198.
2. Quinn McNemar, Psychological Statistics (New York: John Wiley and Sons, Incorporated, 1955), p. 147.
3. Garrett, loc. cit.
4. McNemar, op. cit., p. 385.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November, 1960)

A LONGITUDINAL STUDY OF THE RELATIONSHIP OF HIGH SCHOOL FOREIGN LANGUAGE AND MATHEMATICS STUDY TO FRESHMAN GRADES

JOSEPH V. WEST, Baylor University
BENJAMIN FRUCHTER, The University of Texas

Background

SOME FORTY years ago, Thorndike (8: 421 - 24) convinced himself and many others that formal discipline did not exist. The controversy has subsided somewhat, but there are still many educators who believe that the study of the more difficult subjects, such as mathematics and a foreign language trains the mind for greater life achievement.

Recently Rockwell (3:1-3) gathered new data and also reinterpreted data compiled by Rivenberg (3:1) and Ross (5:20-22) to show that college freshmen's grades were directly related to the amount of foreign language studied in high school and were nearly as closely related to the amount of mathematics studied. The amounts of other subjects studied in high school had either a negative or no relationship with college grades. Rockwell concluded that "... secondary-school study of foreign lamguage improves freshman college grades in all subjects..." (4:306).

An investigation of the relation of grades of the 1951 freshman class in the College of Arts and Sciences at the University of Texas to their secondary-school study of various subjects showed results (9:1) similar to those reported by Rockwell, Rivenburg, and Ross, as noted above. The consistent agreement of the several studies pointed up the need of further investigations to explore some of the underlying relationships. It should be noted that Ross' conclusion from his own data was that "... students are not generally excellent because they take language, but they take language because they are excellent." (5:22)

In attempts to eliminate some possible uncontrolled sources of differences in previous studies, MacPhail (2), Cole (1), Sorenson (7), and Smith and Douglass (6) formed matched groups on the basis of results on tests of scholastic aptitude in studying the effect of different amounts of one foreign language (Latin) on college grades. Sorenson was the only one of the above investigators whose conclusions appeared to support the doctrine of formal discipline; his conclusion was that in predicting academic success, "... Latin seems to differentiate better than do mental test quarters." (7:242). He had matched his groups on the basis of quarters of mental tests rather than exact scores. MacPhail (2:261-62) matched twenty-seven pairs from 485 students and concluded that, since the grades for the Latin students were only three per cent higher and this difference was 2.34 times its probable error, definite conclusions could not be drawn. Cole's results (1:618-22) were similar to those of MacPhail. Smith and Douglass (6:695-701) matched 101 pairs and obtained slight superiority of college grades for students who had studied Latin in high school, but the t-ratio of 1.78 was not significant.

Procedure

The matched groups formed in the above-mentioned studies did not control such factors as socio-economic status and level of performance in school as indicated by high school grades which might contribute to observed differences. To obtain a group whose high school grades were based on a relatively consistent grading system, students from the high school which had the highest proportion of the graduating class enrolled in the College of Arts and Sciences of the University of Texas were selected as subjects for the study. Furthermore, in previous studies scholastic aptitude had been determined at college entrance, after the students had completed high school study of Latin and mathematics. The study of these subjects possibly had an effect on performance on measures of scholastic aptitude taken after such study.

The group of students selected for this investigation had been in one school system since the fourth or fifth grade, at which time each had taken the Elementary Series of the California Test of Mental Maturity. Scores on the 1941 edition of the American Council on Education Psychological Examination were available on all these students at college entrance. Only individuals who were sixteen through nineteen years of age at college entrance and who had completed twelve or more semester hours during their first semester of college attendance were included. Cumulative records indicated that English was the major language spoken in students' homes.

By using the 1950, 1951, and 1952 classes of freshmen, samples of fifty-six males and sixty females were obtained. There were no consistent mean differences among the three classes, but there were appreciable sex differences on nine of the measures. The males were higher on the average number of semesters of high school mathematics, while the females had higher means on California Test of Mental Maturity, Language I. Q., secondary school grade point index, American Council on Education Psychological Examination linguistic score, grade point ratio total, grade point ratio without foreign language, grade point ratio without mathematics, number of semesters of high school foreign language, and number of semesters of high school foreign language plus mathematics.

The following is a list of the variables used in the study, each preceded by its abbreviation as used in the tables:

CMM L--Language I. Q. from the Elementary series of the California Test of Mental Maturity.

CMM NL--Non-Language I. Q. from the same test.

CMM T--Total I. Q. from the same test.

GPI--Grade point index based on secondary school grades. This is the sum of junior and senior high school grades obtained as follows: 1) junior high school grade averages were quantified as A+ = 9, A = 8, . . . , C- = 1; plus 2) quarter of senior high school graduating class quantified as 8 for highest quarter on down to 2 for lowest quarter.

SEI--Socio-economic index based on Warner's (10) quantification of occupation of father and / or mother, weighted by a factor of 7, plus quantification of major source of income, weighted by a factor of 5.

ACPE Q--American Council on Education Psychological Examination, quantitative score on 1941 college edition.

ACPE L--Linguistic score from the same test.

ACPE T--Total score from the same test.

GPR T--Grade point ratio in all courses completed during the first semester of college.

GPR FL--Grade point ratio in one or more foreign language courses completed during the same period. This measure was available on thirty-one males and on thirty-nine females who took foreign language during the first semester of college. Nearly half of these students took a second, third, or fourth semester foreign language course because of high scores made on language placement examinations at college entrance.

GPR w/o FL--Grade point ratio in all courses other than foreign language courses completed during this period.

GPR M--Grade point ratio in one or more mathematics courses completed during this period. Such a measure was available for the forty-two males and for the thirty-nine females who took mathematics during the first semester of college.

GPR w/o M--Grade point ratio in all courses other than mathematics courses completed during this period.

SHS FL--The number of semesters of any foreign language successfully completed in high school.

SHS M--The number of semesters of algebra, geometry, and/or trigonometry successfully completed in high school.

SHS FL + M--The sum of the above two measures.

Grade point index, socio-economic index, and language I. Q. on the California Test of Mental Maturity were found to be significantly correlated with total grade point ratio for separate male and female samples. The first three measures were used as matching variables for the comparison groups mentioned below. It was assumed that if groups were not significantly different in grade point index, socio-economic index, and language I. Q. on the California Test of Mental Maturity, but were widely different on number of semesters of high school foreign language or of high school mathematics, differences in freshmen grades would be more clearly associated with differences in the number of semesters of high school foreign language or the number of semesters of high school mathematics.

The matched groups technique was used to get pairs of sub-groups or individuals as nearly alike as possible on California Test of Mental Maturity Language I. Q., socio-economic index, grade point index, and sex. The difference in each pair of sub-groups was controlled by means of the number of semesters of high school foreign language, the number of semesters of high school mathematics, or the number of semesters of high school foreign language and mathematics taken.

To secure the largest number of individuals in each comparison, the number of semesters nearest to the mean of each controlled variable was used to divide the group for matching purposes. This dichotomization gave results which permitted the comparison of the influence of above-average with below-average number of semesters of high school foreign language, of high school mathematics, or of high school foreign language plus mathematics. Since twenty-nine of the males, but only six of the females, had not taken foreign language in high school, a separate comparison was made to determine the effects of some study versus no study of foreign language in high school for a male sample only.

Results

A summary of the comparisons for the matched groups is presented in Tables I and II. A noteworthy observation on the results recorded in Table I is

TABLE I

SUMMARY OF EIGHT MATCHED GROUP COMPARISONS SHOWING BASES FOR SELECTING SUB-GROUPS AND MEAN AMOUNT OF FOREIGN LANGUAGE AND MATHEMATICS STUDIED IN HIGH SCHOOL

Comparison Number	Number of pairs: Males	Females	Total	Source of Individuals	Basis for selecting Sub-group A	Sub-group B	Mean of SHS FL for Sub-group A	Sub-group B	Means of SHS M for Sub-group A	Sub-group B
I	6	9	15	Total group	≥4 SHS FL	≤3 SHS FL	5.33	0.73	5.80	5.60
II	5	6	11	Those with college foreign language	≥4 SHS FL	≤3 SHS FL	5.27	0.73	6.27	6.00
III	6	3	9	Those with college mathematics	≥4 SHS FL	≤3 SHS FL	4.89	0.44	6.11	5.89
IV	11	6	17	Total Group	≥9 SHS FL+M	≤8 SHS FL+M	5.06	0.53	6.53	5.89
V	5	10	15	Total Group	≥6 SHS M	≤5 SHS M	3.33	4.33	6.47	4.73
VI	4	5	9	Those with college foreign language	≥6 SHS M	≤5 SHS M	2.89	4.44	6.67	4.78
VII	3	5	8	Those with college mathematics	≥6 SHS M	≤5 SHS M	2.50	4.75	6.50	4.63
VIII	16	0	16	Total group	≥1 SHS FL	0 SHS FL	4.06	0.00	6.31	6.31

TABLE II

SUMMARY OF EIGHT MATCHED GROUP COMPARISONS SHOWING MEANS AND DIFFERENCES OF MEANS FOR FIRST SEMESTER COLLEGE GRADES AND FRESHMAN TEST VARIABLES

Comparison	Sub-group	Mean difference for the matched sub-groups on the following variables:						
		GPR T	GPR FL	GPR w/o FL	GPR M	GPR w/o M	ACPE L	ACPE Q
I	A	2.30		2.29			64.80	
	B	2.15		2.21			67.53	
	Diff.	0.15		0.08			-2.73	
II	A	2.85	2.91	2.78			74.91	
	B	2.12	1.73	2.26			75.91	
	Diff.	0.73	1.18	0.52			-1.00	
III	A	2.18			2.33	2.12	68.33	38.33
	B	1.96			1.66	2.01	62.66	47.56
	Diff.	0.22			0.67	0.11	5.67	-9.22
IV	A	2.80		2.72		2.73	69.71	48.23
	B	2.39		2.48		2.43	69.53	45.88
	Diff.	0.41		0.24		0.30	0.18	2.35
V	A	2.59				2.60		46.13
	B	2.34				2.33		42.33
	Diff.	0.25				0.27		3.80
VI	A	2.30	2.22	2.32			66.00	46.44
	B	2.17	2.00	2.22			69.33	45.77
	Diff.	0.13	0.22	0.10			-3.33	0.67
VII	A	2.66			2.88	2.60	71.00	48.25
	B	2.26			2.38	2.22	71.75	44.37
	Diff.	0.40			0.50	0.38	-0.75	3.88
VIII	A	2.24		2.23			63.56	
	B	2.19		2.27			65.12	
	Diff.	0.05		-0.04			-1.56	

that the groups controlled with respect to number of semesters of high school foreign language had nearly the same means on number of semesters of high school mathematics for the two sub-groups, but in the comparisons controlled on number of semesters of high school mathematics, the sub-groups taking more mathematics had from 1.00 to 2.25 fewer semesters of high school foreign language. Since comparison IV (Table I) resulted in little mean difference in number of semesters of high school mathematics but in a large mean difference in number of semesters of high school foreign language, it was regarded as similar to comparison I rather than representing differences in both the number of semesters of high school foreign language and the number of semesters of high school mathematics. The mean differences of the sub-groups on the matching variables for all eight comparisons were kept less than 1) one I.Q. point on the California Language I.Q., 2) one unit on socio-economic index, and 3) one half unit on grade point index.

In spite of the small size of the matched group in comparison II (Table II), the grade point ratio and the grade point ratio in foreign language were appreciably higher for the sub-group which had taken the greater number of semesters of high school foreign language, but the grade point ratio without foreign language difference, although in favor of the group with more semesters of high school foreign language, was not large.

Comparisons III and VII had relatively high mean differences in favor of those who had more semesters of high school foreign language and of high school mathematics, but the grades had such large variability that the mean differences for the grade point ratio in mathematics were small by comparison. In all comparisons the trend was for the differences in the total grade point ratio to be greater than for the grade point ratio without foreign language, or for the grade point ratio without mathematics except in comparison V. These trends indicate that much of the advantage gained by increased study of foreign language and mathematics in high school is carried over to better grades in the same subject areas in the first semester of college. There were no consistent trends for the mean differences on language scores and quantitative scores on the American Council on Education Psychological Examination.

Unique Features and Limitations of the Study

The present study had a number of unique features and limitations which should be stated explicitly and taken into account in interpreting the results.

1) To our knowledge, no previous study on this subject has attempted to make a longitudinal study of individuals from the fourth or fifth grade through the first semester of college.

2) There were appreciable sex differences on a number of the variables, but they were controlled by matching males with males and females with females.

3) The number of pairs available for the various comparisons was considerably reduced from the original 114 students due to the requirements of the matching on the four variables.

4) No control was exercised as to the exact foreign languages involved, but the trends of the relationships were indicative of some transfer from any high school foreign language to any college foreign language.

Summary and Conclusions

Increased scholastic achievement in the freshman year of college had been shown in previous studies to be associated with increased study of foreign language and mathematics in high school The procedure used here to investigate certain aspects of freshman college grades was to form sub-groups matched so that their backgrounds and abilities were nearly alike except for the amount of foreign language and/or mathematics studied in secondary school.

While there was a consistent trend of increased scholastic achievement in all first semester college grades (with one slight exception), the only significant differences of means were for the sub-groups which had taken foreign language in high school and also studied foreign language in college. This difference of more than one letter grade superiority in college foreign language for the sub-group which had taken above-average number of semesters of high school foreign language was reflected in significantly higher grade point average, but the significance was not maintained when the college grade average did not include foreign language grades. The increased study of mathematics in high school showed a consistent trend of increased scholastic achievement in all first semester college grades, and mainly in mathematics grades, but none of mean differences for these comparisons were significant.

REFERENCES

1. Cole, L. E., "Latin as a Preparation for French and Spanish," School and Society, 19: 618-22, May 24, 1924.
2. MacPhail, Andrew R., "Freshman Academic Achievement in College Students Presenting Four Years of Latin and Those Presenting No Latin," School and Society, 19:261-62, March 1, 1924.
3. Rockwell, Leo L., High School Foreign Language Study and College Freshman Grades, mimeographed sheets, Colgate University, Hamilton, New York, no date.
4. Rockwell, Leo L., "Modern Languages in General Education," School and Society, 71: 305-308, May 20, 1950.
5. Ross, C. F., "A Method of Forecasting College Success," School and Society, 34:20-22,

July 4, 1931.

6. Smith, Mary Elizabeth, and Douglass, Harl R., "The Relation of High-School Latin to Marks in the First Year Arts College," The School Review, 45:695-701, November, 1937.
7. Sorenson, Herbert, "High School Subjects as Conditioners of College Success: Implications and Theories Concerning Mental Factors and Faculties," Journal of Educational Research, 19:237-54, April, 1929.
8. Thorndike, Edward L., The Psychology of Learning, Educational Psychology, Volume II, New York: Teachers College, Columbia University, 1913.
9. Unpublished study made by the Testing and Guidance Bureau, the University of Texas, Austin, 1953.
10. Warner, W. Lloyd; Meeker, Marchia, and Eells, Kenneth, Social Class in America, Chicago: Science Research Associates, Inc., 1949.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

INCIDENTAL LEARNING IN FOREIGN LANGUAGE LEARNING

PAUL PIMSLEUR
University of California at Los Angeles

INCIDENTAL LEARNING has been defined as "learning which apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in question."[1]* For example, when a student "just happens to remember" a foreign word which had been present in some reading material the class had read cursorily, then we may call his learning "incidental". The difference between an excellent student and his classmates often lies in the former's added receptivity, his tendency to "pick up" information beyond what is required of him. Such learning is incidental to the extent that he has not deliberately set himself the task of learning, but simply seems to absorb extra information present in, but not essential to, the material he is studying. The purpose of this experiment is to establish 1) whether incidental learning takes place in a language learning situation, and 2) if so, to what extent it takes place.

The literature of psychology abounds in experiments on incidental (or latent) learning. Many of these have dealt with animal subjects;[2] a certain number have used human subjects;[3] These have demonstrated repeatedly that incidental learning does take place, and have attempted to detect some of the relevant factors, such as the kind of instructions the subject is given, the speed of presentation of material, and the effects of reward. It has been found that verbal skills lend themselves more readily than non-verbal ones to incidental learning,[4] and that incidental learning is greatest at the beginning or end of a task.[5] A complete review of experimental findings would greatly increase the length of this paper and would be somewhat peripheral to the subject at hand. This paper confines itself to the experiment it describes, with its most immediate results and conclusions.

Matching of Subjects

The subjects were students in two beginning French classes at UCLA taught by the same teacher at noon and 1:00 p. m., respectively. The two classes were considered to form a single pool of subjects. The teacher was asked to match pairs of students on the basis of their performance in French. Since the experiment took place at about mid-semester, a number of test and quiz marks were available for each student. These marks were independent of the experiment, as the teacher did not know until the time of the experiment that he would be asked to match his students. As finally constituted, each group contained two A students, six B students, five D students and no F Students.

Though matching was done mainly on level of performance, account was taken whenever possible of such other variables as age and sex. The two groups were quite similar in these respects. In view of the procedures employed it is believed that any results obtained can be generalized to all language classes of roughly similar make-up.

The Experimental Stimulus

At the time of the experiment, the students had already learned that there exists in French a tense called the "compound past," and how to form it. They knew the conjugations of the two auxiliary verbs (avoir and être), and they knew how to form the past participle of first conjugation (-er) verbs. They had learned few, if any, irregular past participles, though they knew that not all verbs are regular in this respect. With this stage of development in mind, a grammar drill (the experimental stimulus) was given to the Experimental group, but not to the Control group. The drill was administered in the French Language Laboratory. The experiment was conducted during regular class hours at 12 and 1 o'clock; care was taken to use the identical procedure both times. The students were accustomed to such laboratory sessions and to drills of the kind to be described. They were unaware that an experiment was being conducted until after they had received the experimental stimulus (the drill).

The grammar drill offered practice in the correct placement of pronouns in compound past sentences. In such sentences as Je le lui ai donné, word order is a frequent source of trouble to students. They were given an oral drill in which a noun was to be replaced by a pronoun (e. g., TAPE: Jacques a mangé les legumes; STUDENT: Jacques les a mangés). The drill lasted ten minutes and twenty seconds, during which time 55 sentences were presented, utilizing 24 different verbs. The sentences became somewhat more difficult as the drill progressed, including some negative sentences. In this way, the attempt was made to keep the drill sufficiently taxing so that the students would have to concentrate fully on the required pronoun changes.

---

* All footnotes will be found at end of article

Thus, the ostensible task, as far as any of the students could tell, was to learn the position of pronouns. In actuality, the experimenter was interested in testing their "incidental learning" of the past participles contained in the sentences. After beginning the drill with sentences using 4 familiar verbs, the remainder of the sentences used 20 common verbs whose past participles were unknown to the students. The students had seen all 20 of these at some time during the semester, at least in their infinitive form, but they had not learned their past participles and in most cases had never seen or heard them. While the Experimental group was doing the drill, the Control group was given an unrelated task, a pronunciation exercise. Thus, if the E's know past participles better than the C's, the result can only be ascribed to the influence of the drill, and "incidental learning" can be said to have taken place.

The Testing Procedure

As has been explained, it was desired to test students' knowledge of past participles. Two kinds of knowledge had to be tested: The lesser (passive) kind called by psychologists "recognition," and the greater (active) kind termed "recall." A further consideration: since the drill was oral, the testing should be oral, but a written test was also desirable to see if information acquired by "incidental learning" would carry over to written work. Therefore, two tests were constructed, an oral and a written one. Each of these was further split into two halves, one half to test recognition and the other to test recall. Thus the four sub-tests were: oral recognition, oral recall, written recognition, written recall. Combinations of these sub-tests would provide such scores as: 1) Total Oral, and 2) Total Written (recognition and recall); 3) Total Recognition and 4) Total Recall (oral and written); 5) Grand Total (all four sub-tests combined). In this manner the opportunity existed for probing the extent of "incidental learning," should it prove to take place at all. Test reliabilities are shown in Table I.

Regarding the validity of the oral and written tests, no evidence is offered other than the manifest connection between the content of the test and the trait being tested. Since only surface behavior is being tested, on the theory that all the student "really knows" about past participles will be demonstrated by his test answers, face validity appears sufficiently convincing in this instance.

Analysis of Data; Results

The first phase of the statistical analysis of test scores was designed to find differences in several variables simultaneously. One variable is group membership; does the Experimental group differ from the Control group in their scores on the tests? Another variable concerns level of performance; do the A, B, C and D students differ? A further question concerns possible interaction between these two variables; does the combination of group membership (E or C) with level of performance make a difference; specifically, did the high level (A and B) students in the Experimental group profit more than expected from the drill to which they were exposed? The analysis of variance method[6] was used to test all of these aspects at the same time.

First to be tested were the most general data, the Grand total scores (the total of scores on all four sub-tests). Bartlett's test for homogeneity of variance[7] resulted in a non-significant chi-square between rows. The analysis of variance showed a significant difference in Grand Total scores between students of different levels of performance. This is no more than a confirmation of the fact that A, B, C and D students differ. The data fail to reveal a difference between the Experimental and Control groups, or any interaction between group and level.

The Grand total scores upon which this analysis was performed are the least sensitive of the scores. That is, they lump together the factors of oral vs. written work and of recognition vs. recall. If any significant difference exists between the Experimental and Control groups, it should be present in the Oral Recognition score. The results of the analysis of Oral Recognition scores are shown in Table II.

As will be seen from the data in Table II, there are two significant findings. The difference in level of performance is once again confirmed, at the 99% level of confidence. The other finding interests us more, for the Experimental group is found (with 95% confidence) to be superior to the Control group in Oral Recognition of past participles. With this finding, incidental learning is shown to have taken place. However, that its effects are limited can be inferred from the presence of a difference between the E and C groups in the "easiest" test and the absence of such a difference in the "hardest" test. Thus incidental learning seems to take place to a moderate degree; how moderate the degree is to be determined by further analysis of the data. The method of analysis of variance used thus far is abandoned in favor of t-tests of differences between the E and C groups on various part scores. The t-tests answer the one question still remaining (in which sub-tests do the groups differ?), now that it has been made apparent that no interaction will be found and that a difference among levels of performance will always be found. The t-tests provide a

TABLE I

TEST RELIABILITIES, SPLIT-HALF, CORRECTED FOR LENGTH.

	Recognition	Recall	Total
Written	.57	.73	.71
Oral	.89	.76	.76
Total	.81	.61	Grand .94

TABLE II

ANALYSIS OF VARIANCE ON ORAL RECOGNITION SCORES

Source	SS	df	MS	F
Between cols. (E and C gps)	22	1	22	5.23*
Between rows (A, B, C, D levels)	86	3	28.67	6.81**
Interaction	10	3	3.33	not sig.
Within cells	118	28	4.21	-------
	236	35		

*Significant .05
**Significant .01

TABLE I

MEANS, VARIANCES, AND t VALUES FOR 59 OVERACHIEVERS AND 44 UNDERACHIEVERS ON CIAA PART AND TOTAL SCORES

Part	Overachievers		Underachievers		t value
	$\overline{X}$	s^2	$\overline{X}$	s^2	
I	17. 63	15. 03	15. 18	18. 65	2. 97**
II	21. 83	20. 90	18. 79	22. 95	3. 28**
III	22. 93	98. 14	17. 93	36. 44	3. 17**
IV	27. 25	62. 50	27. 34	41. 26	0. 06
V	20. 63	31. 14	19. 00	45. 53	1. 30
VI	18. 63	13. 66	16. 70	18. 53	2. 38*
Total	128. 93	575. 17	115. 00	549. 58	2. 95**

* Significant beyond the . 05 level of confidence.
**Significant beyond the . 01 level of confidence.

those subjects who failed to respond to these mailed requests. Finally, those students who had not responded to the mailed requests were contacted personally and urged to fill out the inventory. As a result of this procedure, all 103 subjects completed the CIAA.

Analysis

The initial analysis involved testing the significance of difference between means for men and women. Since no significant differences were found, both men and women were included in the final analysis.

The significance of differences between means for overachievers and underachievers on the six part scores and total score of the CIAA was assessed by t tests.

Results

With the exception of Part IV, the overachievers' means were consistently larger than the means for underachievers. The differences between means for Parts I, II, III, and total score were significant beyond the .01 level of confidence. The difference between means for Part VI was significant beyond the .05 level of confidence. Differences between means for Parts IV and V were not significant. Means, variances, and t values for the 59 overachievers and 44 underachievers on CIAA part and total scores are presented in Table I. The considerable overlap between overachievers and underachievers for all parts and the total score should be noted.

Summary Statement

This study was designed to further assess the validity of the CIAA by comparing the scores of academic overachievers and underachievers on the six part scores and total score. From 405 college sophomores, 59 students were identified as overachievers and 44 students were identified as underachievers. The CIAA was personally administered to all of the 103 subjects. The means of the overachievers on four part scores and the total score were significantly larger than the means of the underachievers. It was concluded that the CIAA may validly be employed to discriminate between academic overachievers and underachievers at the college level.

Results of this study are in general agreement with those of the Christensen investigation, namely, significant differences were found between means for overachievers and underachievers on Parts I, II, III, and total CIAA scores. In the present study a significant difference on Part VI, also in favor of the overachievers, was noted. The present study supports the validity of the CIAA as a device which may be employed to discriminate between academic overachievers and underachievers at the college level. The value of the CIAA to counselors was not considered in this investigation.

REFERENCES

1. Borow, H. A Psychometric Study of Non-Intellectual Factors in College Achievement. Unpublished Ph.D. Dissertation, Pennsylvania State College, 1945.
2. Borow, H. Manual for the College Inventory of Academic Adjustment. Stanford: Stanford University Press.
3. Christensen, C. M. "A Note on Borow's Inventory of Academic Adjustment, Journal of Educational Research, L (September, 1956), pp. 55-58.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 3, November 1960)

FACULTY AND STUDENT PERCEPTIONS OF CHEATING

JACK R. FRYMIER
Director of Instruction
Orlando, Florida

The Problem

WHAT IS cheating? Does a university faculty and its student body see cheating in the same way? During the academic year 1957-58, the problem of cheating at a large Eastern university assumed such proportions that student leaders requested an investigation by the administration. The faculty Senate promptly designated a committee to work on the problem, but at the beginning of the fall term, 1958, the problem still loomed. The article below appeared as a student committee report in the student newspaper, and a major news story appeared in The Philadelphia Inquirer the following day regarding these events:

> Cheating is almost totally accepted practice at the University. Since it is apparent that any exam which is re-used will be accessible, no examinations should ever be re-used. Exams should be varied distinctly from semester to semester. It seems that in some cases new examinations leak out to students. This may call for tighter supervision in the makeup and duplicating of exams. All too often proctoring is a farce. More attentive and alert proctors should be utilized to discourage any cheating during the examination itself. The action taken by the Academic Discipline Committee should be publicized to make students aware of the penalties incurred through cheating.

In this setting the author undertook to ascertain and then compare faculty members' and students' perceptions of cheating.

Anderson* described the development and use of an instrument for determining students' perceptions of cheating. Twenty-eight specific behaviors were described, and each of these was seen in varying degrees as cheating or non-cheating by university students.

The Procedure

A 50 per cent random sample of the faculty of the university's three undergraduate colleges (Business, Liberal Arts, and Teachers) were mailed a copy of Anderson's instrument. An accompanying letter discussed the then-current problem on campus of cheating, and the faculty and student committees at work. A total of 177 questionnaires were sent: 94 to members of the Liberal Arts faculty, 45 to members of the Teachers College staff, and 38 to staff members in the School of Business and Public Administration. Of this number, 111 were returned, a 63 per cent response.

Four university instructors administered the same instrument to members of their classes. In all, 130 students completed the questionnaire. Approximately one third of these students were from the Liberal Arts College, about one tenth from the School of Business, and the remainder from Teachers College. All responses were made anonymously.

Anderson's instructions and scoring procedure (rating items 1 to 5) were followed as described, a score of 5 meaning that the particular behavior of a fictitious student in a fictitious situation was "entirely justified," and a score of 1 indicating that the behavior was "definitely not justified."

For both groups total score means and item score means were computed. The statistical significance of the difference between total score means was determined by computing $\underline{t}$. Further, the number of responses which both the faculty and students' group made which were "extreme," i.e., marked either 1 or 5, was also determined, and the

---

* William F. Anderson, Jr. "Attitudes of University Students Toward Cheating," Journal of Educational Research, 50: 581-8.

TABLE I

COMPARISON OF FACULTY AND STUDENT PERCEPTIONS OF CHEATING

	N	Range	Mean	S. D.	t
Faculty	111	34-72	51.35	6.90	5.17
Student	130	34-104	58.10	12.06	

TABLE II

MEAN SCORES ON VARIOUS ITEMS OF CHEATING QUESTIONNAIRE BY FACULTY AND STUDENT GROUPS

Faculty Mean Score	Behavior Described In Item	Student Mean Score
4.91	Studying with classmates before law exam	4.91
4.65	Studying old tests from sorority file	3.55
4.36	Sorority hiring graduate student to tutor	4.34
4.29	Talking to instructor after class	4.42
3.76	Getting help from friend who had completed course on kind of questions usually asked	3.89
2.73	Telling instructor she enjoyed the course	2.65
2.68	Asking friend who had just completed test previous hour about questions asked	2.99
1.85	Writing down formula when first enter room so will not forget	2.05
1.78	Using same term peper for two courses	2.91
1.33	Going thru waste cans and finding copy of test	1.69
1.18	Borrowing roommate's term paper and turning it in after a few minor changes	1.53
1.15	Looking at neighbor's paper during exam	1.55
1.14	Helping another student who is having trouble during an exam	1.61
1.13	Hiring English major to write term paper	1.60
1.11	Using notes after being put on "honor"	1.32
1.07	Looking at notebook when instructor is not looking	1.21
1.06	Putting notes on adhesive tape on inside of skirt hem	1.21
1.05	Looking at text while instructor writes questions on blackboard	1.66
1.04	Getting copy of exam from friend who is a secretary	1.39
1.04	Looking at kleenex with notes on it	1.21
1.04	Using notes hidden in hand during exam	1.20
1.03	Paying grader $10 to give "B" on paper	1.34
1.03	Coughing during exam to communicate to friend correct answers on true-false test	1.30
1.03	Making scroll of match sticks and paper for use during exam	1.21
1.03	Leaving room during exam to check notes in rest room	1.19
1.01	Asking neighbor for answers during exam	1.22
1.01	Paying friend to take final exam in large class	1.16
1.00	Turning in empty blue book, then going into instructor's office and changing books	1.15

difference between the two groups on this factor determined by ohi square.

The Results

Table I shows the total score comparison of the faculty and student groups. Inasmuch as the instrument consisted of 28 items, a minimum score of 28 and a maximum score of 140 were possible.

University faculty members' responses ranged from a low of 34 to a high of 72. University students' total scores, on the other hand, ranged from a low of 34 to a high of 104. This difference is demonstrated further by comparing the faculty group's total score mean of 51. 35 to the student group's total score mean of 58. 10. Statistically, this difference is significant at the . 001 level of confidence.

The specific differences between faculty and students' perceptions of cheating may be studied by examining Table II. In 24 of the 28 specific behavioral situations rated faculty members were more severe in labeling the incident "not justified" than were the students. Many of these differences, however, were very slight. In one instance both groups rated the item exactly the same, and in three the faculty members were less severe than were the students in condemning the particular behavior. Although the statistical significance of the difference between faculty and student group response for each individual item was not computed, only about three items seem to differ very much.

During the course of the study it became apparent that faculty members tended to categorize many behaviors as "entirely justified" or "definitely not justified" by employing the extremes of the rating scale-- 1 or 5 -- much more frequently than the middle scores. Almost 87 per cent of the faculty's responses were either 1 or 5, as compared to 73 per cent of the students' responses falling in the same categories. When ohi square was computed to determine if the actual frequencies differed statistically, a figure of 187. 49 was obtained. This was significant beyond the . 001 level of confidence.

Discussion of Results

These data seem to indicate that faculty members and university students see cheating differently. In a few instances these differences seem to be specific, i. e., regarding particular behaviors, but in more cases they seem to be generalized. Faculty and students actually seem to see various examples of student behavior similarly in direction but differently in degree. And, as might be expected, it is the faculty members who generally seem to be more severe in describing a particular behavioral act as cheating than are the students.

All things considered, then, even though the statistical differences are considerable, there seems to be very little basic difference in the orientation of these two groups toward cheating. The overall pattern of perceiving various behaviors as cheating or non-cheating seems to be quite similar, and this pattern of similarity is probably more significant than the statistical differences which exist.

Both groups seem to agree that getting help before an examination is usually justifiable, but getting assistance during the exam or by any covert means is inappropriate.

It may be that the instrument used for assessing these person's perceptions of cheating is too crude. Very few fine differentiations seemed to occur, and this certainly could be an indication of the instrument's ineffectiveness. Refinements of some of the items might well add to its utility. An exploration of some of the more subtle aspects of cheating, also, might well lead to worthwhile results. For example, if an instructor grades on a curve and makes no effort to curb cheating during an examination, is such behavior justifiable? Further research could prove quite helpful.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 DECEMBER 1960 NUMBER 4

CONTENTS

One Explanation of the Extent of Function Fluctuation Observed in Total and Subtest Cognitive Performance—C. C. Anderson ... 123

The Alternative Course in Secondary Mathematics—Daniel S. Parkinson ... 128

Toward a Theory of Interpersonal Relationships in the Administration of Higher Education—David L. McKenna ... 133

Selected Out-of-School Factors That Affect Negro Elementary School Children—Elsie Hill Wallace ... 137

Missing Data in Non-Linear Trend Analysis of Repeated Measurements on the Same Individuals—Richard B. Seymour ... 141

Special Budgetary Provisions Made for Educating the Gifted—Morris J. Applbaum ... 145

An Attention Scale for Evaluating E. T. V. Programs—Cecil C. Bridges, Jr. ... 149

The Effects of the Purported Validity of a Test Upon Level of Motivation and Performance—Lawrence S. Wrightsman, Jr. ... 153

Intercorrelations of the American Council Psychological Examination with Other Types of Tests—J. C. Gowan ... 157

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

ONE EXPLANATION OF THE EXTENT OF FUNCTION FLUCTUATION OBSERVED IN TOTAL AND SUBTEST COGNITIVE PERFORMANCE*

C. C. ANDERSON
University of Alberta

ANDERSON (1, 2) HAS found, generally, that performance by English children on a test of verbal intelligence showed, on the average, significant evidence of fluctuation and, in particular, that the extent of fluctuation characteristic of the functions measured by the subtests in a non-verbal test of intelligence followed a trend, the earliest and latest subtests showing the greatest amounts of fluctuation, the middle tests the least. If one of Anderson's hypotheses is true—that this trend might be a product of the position of the subtests in the series— then clearly a similar trend would appear if another cognitive test were used. Accordingly the present paper is an attempt to find evidence of this trend in the subtests of the Terman-McNemar Test of Mental Ability, and to establish whether the performance of Canadian children on the total test showed an amount of fluctuation approximately similar to that of the English children. A very subsidiary aim was to examine the significance of sex differences in fluctuation.

Procedure

Forms C and D of the Terman McNemar Test of Mental Ability were administered together on each of four different testing occasions, separated from each other by an interval of one week, to two classes of pupils in grade eight. On the initial testing with Form C, the mean I. Q. was 100, the standard deviation being approximately 12. The test for evidence (criterion) and measure of the amount (group index) of function fluctuation were those originally suggested by Thouless (3) and recently used by Anderson. The rationale behind Thouless's basic statistic (the criterion) is that evidence of fluctuation appears when changes in scoring between any two administrations of one test are correlated both in amount and direction with changes in scoring between the two administrations of the alternative prallel form.

For example in the present case, if fluctuation between the first two testing occasions is being considered, the criterion will be $r(C_1 - C_2)(D_1 - D_2)$, which is usually calculated in the form:

$$r_{C_1} + r_{C_2 D_2} - r_{C_1 D_2} - r_{D_1 C_2} /2 \sqrt{(1 - r_{C_1 C_2})(1 - r_{D_1 D_2})}.$$

Since the criterion is partly determined by the size of the tests' reliability coefficients, an accurate measure of the amount (group index) of fluctuation can be had by dividing this criterion by the average of the two within—occasion reliability coefficients obtained from any combination of two testing occasions--in this case $r_{C_1 D_1}$ and $r_{C_2 D_2}$. Because there were four testing occasions and therefore six possible pairs of testing occasions (1 - 2, 1-3, 1 - 4, 2-3, 2-4 and 3-4), six criteria and indices were available.

Results

Analysis of Total Scores. These criteria and indices are presented in Table I.

The sex differences in amount of fluctuation were found to be insignificant.

None of the criteria are significant and the average index of only .122 is significantly smaller than the index derived from the cognitive test performance of children in England of similar age (2). There are two possible major explanations of this finding. Firstly, it might be argued that the Canadian educational culture differed from the English in some at present unknown property which operated to produce this observed decrease in fluctuation. If this is the case, then certainly that unknown property was not differential motivation because, in the present experiment, a considerable effort was made to motivate the children to achieve, partly by offering prizes to anyone making agreed increments of score between successive testing occasions and partly by allowing the children to believe that their

---

* Footnotes will be found at the end of the article.

TABLE I

CRITERIA AND INDICES DERIVED FROM ALL POSSIBLE COMBINATIONS OF TESTS AND RETESTS ON FOUR TESTING OCCASIONS

Combinations of Testing Occasions	Criterion	Index
1 - 2	.170	.184
1 - 3	.120	.130
1 - 4	.124	.134
2 - 3	.132	.142
2 - 4	.162	.176
3 - 4	-.032	-.035

TABLE II

AMOUNTS OF FLUCTUATION OBSERVED BETWEEN SIX PAIRS OF TESTING OCCASIONS FOR EACH OF THE SEVEN SUBTESTS OF THE TERMAN-McNEMAR TEST OF MENTAL ABILITY

Pairs of Testing Occasions	Subtests						
	1	2	3	4	5	6	7
1 - 2	-.040	-.094	-.018	.179	.093	.279	.090
1 - 3	.188	-.073	.124	.119	.117	.159	.398
1 - 4	.479	-.179	.359	.162	.086	.069	.120
2 - 3	.042	.232	.531	.442	.036	-.195	.367
2 - 4	.490	.077	.393	.296	.319	.222	.435
3 - 4	.218	-.315	.238	.332	-.104	-.043	.444
Average	.230	-.059	.271	.255	.091	.082	.309

TABLE III

ANALYSIS OF VARIANCE OF GROUP INDICES OF FLUCTUATION BETWEEN SIX COMBINATIONS OF TESTING OCCASIONS BY BOYS AND GIRLS IN THE SEVEN SUBTESTS OF THE TERMAN-McNEMAR TEST OF MENTAL ABILITY

Source of Variance	Sum of Squares	df	Mean Square	F
Subtests (ST)	2.284	6	.381	3.073*
Sex (s)	.066	1	.066	.532
Combinations of Testing Occasions (C)	.445	5	.089	.718
ST x S	1.501	6	.250	2.016
ST x C	2.607	30	.086	.694
S x C	.384	5	.077	.621
ST x S x C	3.742	30	.124	
Total	11.029			

* Significant at the 5% level

TABLE IV

TOTAL NUMBER OF WRONG RESPONSES TO EACH SUBTEST MADE BY ALL PUPILS ON ALL OCCASIONS

	Subtests						
	1	2	3	4	5	6	7
Wrong Responses	3,000	5,652	3,674	2,972	2,828	5,222	1,803*

* The maximum number of wrong responses to subtest 7 is only 12 as against 25 for the other subtests and therefore the total 1,808 was arrived at by multiplying the observed total by 2-1/12.

TABLE V

ANALYSIS OF VARIANCE OF THE TOTAL NUMBERS OF WRONG RESPONSES TO THREE SUBTESTS SHOWING MOST EVIDENCE OF FLUCTUATION AND TO THREE SUBTESTS SHOWING LEAST EVIDENCE OF FLUCTUATION

Source of Variance	Sum of Squares	df	Mean Square	F
Between Groups of Subtests	342,732	1	342,732	20.62*
Within Groups	764,544	46	16,621	
Total	1,107,276	47		

* Significant at the 1% level.

TABLE VI

TOTAL NUMBER OF ITEMS UNATTEMPTED BY ALL PUPILS ON ALL OCCASIONS IN RESPONSE TO TWO GROUPS OF THREE SUBTESTS

Subtests	Group 1	Subtests	Group 2
1	17	2	148
3	17	5	12
4	5	6	60

performance on the tests would count toward their final marks for the year.

The second explanation of the different amounts of fluctuation characteristic of the Terman-McNemar and Moray House Test is in terms of the relative extent to which they are speeded. The manual of the former claims that it is a power test, while the present author has the impression that the latter is to a great extent speeded, although the Moray House manual gives no information on this point. If this assumption is correct, then this explanation of cognitive fluctuation would be phrased in the following way: that the extent to which performance on a cognitive test shows evidence of fluctuation is correlated with and determined by the extent to which the test is speeded. In the present writer's experience, clerical tests [2] almost purely of speed have shown the greatest amounts of fluctuation, the Terman-McNemar (a test of power) has shown least fluctuation, and the partly-speeded Moray House and N. F. E. R. tests occupy a median position.

Analysis of Subtest Scores. This hypothesis is substantiated by evidence concerning the amount of fluctuation observed in performance on the subtests of the Terman-McNemar. Scores on the two parallel forms of each of the seven subtests were available for the four testing occasions and the six group indices for each subtests are presented in Table II.

There is clearly no evidence of the sort of trend which Anderson observed before, and yet the variaability among the indices of the subtests is significant, as can be seen from Table III where the extent of the variability of sex difference is also assessed.

The most plausible explanation of the pattern of these differences in fluctuation is that the subtests which show the greatest amount of fluctuation (1, 3, 4 and 7) are relatively speed tests while those showing the least amount of fluctuation are relatively power tests. This hypothesis can be verified by first taking as a measure of the extent to which tests are power tests the number of wrong responses made on them by all pupils on all testing occasions and by then examining the trend of the totals to determine whether subtests 1, 3, 4 and 7 have totals of wrong responses which are significantly smaller than the corresponding totals of subtests 2, 5 and 6. The totals are presented in Table IV.

The pattern of these totals tends to support the hypothesis since the totals for subtests 2, 5 and 6 are greater than the totals for 1, 3, 4 and 7. However the hypothesis as it is applied to the subtests can be completely verified only if all the totals of wrong responses to subtests 2, 5 and 6 are significantly greater than all the totals of wrong responses to subtests 1, 3 and 4[3]. Since each type of subtest was administered eight times (parallel tests worked on four testing occasions), 24 totals of wrong responses were available for each of the two groups of three subtests and the significance of the difference between these two groups in point of total numbers of wrong responses is estimated in Table V. This is clear evidence of an association between the extent to which a subtest is speeded and the amount of fluctuation it shows.

Another Index of Power. Even if it is argued that the total number of wrong responses is not an accurate measure of the extent to which a test is one of power, nevertheless the same association appears when the form of the index is altered to the number of items unattempted by the student. Table VI presents evidence of this.

Summary

From an examination of the performance of 55 grade eight children on the complete Terman-McNemar Test of Mental Ability and on its subtests, it was concluded that the extent to which a test or subtest shows evidence of fluctuation is correlated with and determined by the extent to which it is a test of speed. Such an hypothesis has the merit of being open to easy experimental verification.

FOOTNOTES

1. The author is indebted to the General Research Fund of the University of Alberta for a grant in aid of this research.
2. For example, the Clerical Test (very similar to the D. A. T. Clerical Speed and Accuracy) devised by the National Foundation for Educational Research in England and Wales.
3. It is possible to argue that the tactic of equating subtest 7 to the other subtests in point of maximum total possible wrong responses is unsound and this subtest has been omitted from this analysis despite the fact that it contributes most to the establishment of the hypothesis in that it shows the largest amount of fluctuation and the smallest total of wrong responses.

REFERENCES

1. Anderson, C. C. "The Extent and Factorial Structure of Function Fluctuation in a Mixed Group of Adolescents and Preadolescents," British Journal of Educational Psychology, to appear.
2. Anderson, C. C. "Function Fluctuation," British Journal of Psychology Monograph Supplement, XXX, 1958.
3. Thouless, R. H. "Test Unreliability and Function Fluctuation," British Journal of Psychology, XXVI, (1936), 16-31.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

THE ALTERNATIVE COURSE IN SECONDARY MATHEMATICS

DANIEL S. PARKINSON
College of Wooster
Wooster, Ohio

RECENT CONCERN regarding the curriculum in high school mathematics has focused attention primarily upon courses designed to meet the needs of students with either superior ability or motivation. Little attention has been paid to courses designed for students who have no interest or purpose in following the sequential or "traditional" program in mathematics. That this is unfortunate is shown by the increasing need for students to understand the technical nature of modern society and to develop the level of mathematical competence necessary for the ordinary affairs of life. The purpose of this paper is to examine the subject content of the alternative course to algebra and geometry in Wisconsin high schools, that is general mathematics.

The two track approach to the organization of high school mathematics is not the only one which exists; it is, however, one which has a long and acceptable history. The possibility of correlated or general courses in high school mathematics was first brought forward by Eliakin H. Moore in his presidential address to the American Mathematical Society in 1903 (1), and that such courses existed was recognized by the National Committee on Mathematical Requirements of the Mathematical Association of America in 1923 (2). The only clear-cut proposal that the two track system is the way in which the curriculum should be organized, however, did not come until the second report of the Commission on Post-War Plans of the National Council of Teachers of Mathematics in 1945. The Commission recommended a two-track program, the first track to provide mathematical training for future leaders in the learned fields, and the second to provide mathematical competency in the ordinary affairs of life for all citizens. As the Commission stated, "The large high school should provide in grade nine a double track in mathematics, algebra for some and general mathematics for the rest" (3). In spite of the proposals for a second track in mathematics, however, the nature of the contents of this course is still unknown.

In order to examine the areas of content found in general mathematics, three textbooks in this field were examined and thirty-nine topics were abstracted. These were divided into twenty-three mathematical-process topics, such as common fractions and graphs, and sixteen mathematical-process topics, such as common fractions and graphs, and sixteen mathematical-application topics such as insurance and property taxes. They were then arranged in a check list as a part of a questionnaire for the head of the department of mathematics. As it was desired to study that type of general mathematics which exists as an alternative to algebra, the study was limited to schools offering both algebra and general mathematics whose graduation requirements did not specify either of these courses. A group of 219 Wisconsin high schools offering both these subjects was selected and the questionnaire was sent to the head of the department of mathematics in the Fall of 1958. Of these 198 responded and thirty-two were excluded leaving 166 schools, or 83.8 per cent of the 198 returns, to be included in the study. Because of the criterion regarding graduation requirements it was impossible to tell if a school would have been included in the study until returns from it were examined, and it is therefore only possible to say that the 166 schools represent a percentage of a potential population in Wisconsin which is somewhere between 89.2 and 100.0. These 166 schools are broken down on the basis of whether or not they have a college-preparatory track, 95 having such a track and 71 not having it.

In order to determine the source of the topics most commonly found in general mathematics, the mathematical-process topics were divided into those taken from arithmetic, from algebra, and from geometry and trigonometry. Of the eight topics taken from arithmetic the fundamental operations, common fractions, decimals, ratio and percent, and area measurement were all offered by over ninety-five percent of the schools (Table I). In addition volume measurement was offered by over ninety-three percent. The remaining two topics, square root and the metric system, were offered by 72.9 and 59.6 percent respectively.

Of the seven topics taken from algebra only two, area formulas and volume formulas, were offered by over ninety percent of all the schools. A third topic, solving linear equations in one unknown, was offered by over eighty percent. Of the four remaining topics, positive and negative numbers were offered by almost seventy percent of the schools, the use of exponents by 54.2 percent, laws of levers by 18.7 percent and logarithms by 3.6 percent.

Of the eight topics taken from geometry and trigonometry, only one, straight line measurement, was offered by over ninety percent of the schools. Three topics, the use of compass and protractor, graphs, and reading angles, were offered by over eighty percent. One topic, geometric constructions, was offered by over seventy percent. Of the remaining three topics, making accurate drawings was offered by 59.6 percent of all the schools, indirect measurement (without use of the trigonometric functions) was offered by 44.0 percent, and indirect measurement (using sine and cosine) was offered by 14.5 percent.

Space was provided for additional topics, if the given list did not include all of those in a particular course in general mathematics. Six topics were added by schools with a college preparatory track: casting out nines, elementary statistics, short methods of multiplying and dividing, latitude and longitude, scientific notation, and other systems. Elementary statistics was listed by three schools and short methods of multiplication by two. The rest were named by only one. Two topics were added by schools without a college-preparatory track: elementary statistics and vectors, neither by more than one school. Because of the small number of topics added and the small number of schools adding them, it is possible to assume that the list provided was reasonably complete.

When the three types of topics (those from arithmetic, from algebra, and from geometry and trigonometry) are examined together it is found that a difference exists in the percentages of schools offering them. For example, of the nine topics offered by over ninety percent of the schools, six were taken from arithmetic, two from algebra, and one from geometry and trigonometry. If the percentage offering is examined down to seventy percent, seven of the fifteen topics included are taken from arithmetic, three from algebra, and five from geometry and trigonometry. Because of the higher percentages of schools offering topics from arithmetic, it is possible to say that general mathematics, as it exists as an alternative to algebra in Wisconsin high schools, is weighted with materials from this area.

When the percentages of schools with and without college-preparatory tracks offering these topics were compared it was found that in eighteen of the twenty-three cases a slightly higher percentage of schools with a college-preparatory track offered the topic than of schools without such a track. In only three cases, square root, graphs, and reading angles, was this tendency reversed. In two cases, ratio and percent, and area measurement, the percentages were exactly equal. The differences in the percentage offerings were often small, amounting in some cases to a few tenths of a percent. It is still possible to say, however, that a tendency exists in this group of schools for courses in general mathematics to include a greater variety of topics if they are offered in schools with a college-preparatory track, than if they are offered in schools without this track.

The sixteen mathematical-application topics were intended to represent those areas in which the practical applications of the mathematical-process topics are shown to students. Three of these topics were included in general mathematics by over ninety percent of all the schools: mathematics in the use of money, mathematics in business, and insurance (Table II). Three others which were included by over eighty percent of the schools were banking, buying with time payments, and budgets. Two topics were included by over seventy percent: property taxes and mathematics in transportation. Five other topics were included by between fifty and seventy percent: social security, stocks and bonds, federal income tax, mathematics in health and safety, and reading meter dials. The remaining three topics were mathematics in agriculture (included by 48.2 percent), use of the Vernier scale and micrometer (included by 25.9 percent), and the slide rule (included by 1.2 percent). It can thus be seen that general mathematics, in the schools included in the study, includes a number of areas in which practical applications are attempted, and that these tend to fall into the areas of personal and commercial mathematics.

The alternative course to algebra, as it appears in Wisconsin high schools, thus appears to be a course which contains a large amount of arithmetic, with an introduction to algebra and geometry, and in which an attempt is made to show the practical applications of its contents. It should be noted that the instrument used in getting this information did not provide a means for judging the time and effort expended on a particular topic. It only recorded whether or not a topic was present or absent in a school. The judgment concerning the relative weight of materials in general mathematics therefore is based upon the assumption that all of the checked topics occupied equal space in the course. It is possible that the greater number of topics which tended to be included in general mathematics by schools with a college-preparatory track is due to some different purpose for the course, such as using it as a preparation for later work in algebra. This question might be worthy of further investigation, perhaps through a selected sampling of students.

It was assumed at the beginning that the technical emphasis of modern society has created a need for at least some knowledge of the mathematical processes involved in it on the part of all students, both in order to understand society and to be able to function effectively in it. It is possible that the

TABLE I

PERCENTS OF SCHOOLS INCLUDING CERTAIN MATHEMATICAL PROCESSES IN THEIR COURSE IN GENERAL MATHEMATICS

Mathematical Processes from the Subjects Listed	Percent of Schools With College-Preparatory Track	Percent of Schools Without College-Preparatory Track	Total
	N = 95	N = 71	N = 166
From Arithmetic:			
Fundamental Operations (addition, subtraction, multiplication and division)	98.9	98.6	98.8
Common Fractions	100.0	98.6	99.4
Decimals	100.0	98.6	99.4
Ratio and Percent	100.0	100.0	100.0
Square Root	72.6	73.2	72.9
Area Measurement	95.8	95.8	95.8
Volume Measurement	93.7	93.0	93.4
Metric System of Measurement	60.0	59.2	59.6
From Algebra:			
Area Formulas	96.8	94.4	95.8
Volume Formulas	94.7	94.4	94.6
Solving Linear Equations in One Unknown	81.1	80.3	80.7
Positive and Negative Numbers	71.6	66.2	69.3
Use of Exponents	54.7	53.5	54.2
Logarithms	4.2	2.8	3.6
Laws of Levers	20.0	16.9	18.7
From Geometry and Trigonometry:			
Straight Line Measurement	95.8	93.0	94.6
Use of Compass and Protractor	85.3	80.3	83.1
Making Accurate Drawings	60.0	59.2	59.6
Graphs	86.3	93.0	89.2
Geometric Constructions	76.8	73.2	75.3
Reading Angles	78.9	81.7	80.1
Indirect Measurement (without use of the Trigonometric Functions)	46.3	40.8	44.0
Indirect Measurement (using sine and cosine)	14.7	14.1	14.5

TABLE II

PERCENTS OF SCHOOLS INCLUDING CERTAIN MATHEMATICAL APPLICATIONS IN THEIR COURSE IN GENERAL MATHEMATICS

Mathematical Applications	Percents of Schools		
	With College-Preparatory Track	Without College-Preparatory Track	Total
	N = 95	N = 71	N = 166
Mathematics in Transportation	72.6	74.6	73.5
Mathematics in Health and Safety	50.5	63.4	56.0
Insurance	90.5	90.1	90.4
Mathematics in Business	92.6	93.0	92.8
Stocks and Bonds	63.2	60.6	62.0
Mathematics in the Use of Money	95.8	93.0	94.6
Mathematics in Agriculture	47.4	49.3	48.2
Property Taxes	73.7	83.1	77.7
Slide Rule	2.1	0.0	1.2
Budgets	82.1	78.9	80.7
Banking	86.3	81.7	84.3
Social Security	69.5	67.6	68.7
Use of Vernier Scale and Micrometer	25.3	26.8	25.9
Federal Income Tax	57.9	69.0	62.7
Buying with Time Payments	81.1	83.1	81.9
Reading Meter Dials	53.7	53.5	53.6

arithmetic emphasis found in the present course in general mathematics is sufficient for the ordinary affairs of life. The question may be asked, however, of whether or not it is possible for persons to have a sufficient understanding of the problems of modern life, in essence to be a good citizen, with only a knowledge of arithmetic and some introduction to algebra and geometry. If it is true that the mathematical understanding provided by general mathematics is not sufficient to meet this need, then it would follow that a critical reexamination of this course is in order.

REFERENCES

1. Eliakin H. Moore, "On the Foundations of Mathematics," Science, 17:406-409 (March 1903).
2. National Committee on Mathematical Requirements of the Mathematical Association of America, The Reorganization of Mathematics in Secondary Education, The Association, 1923, p. 13.
3. Commission on Post-War Plans of the National Council of Teachers of Mathematics, "Second Report of the Commission on Post-War Plans," Mathematics Teacher, 45:195-221 (May 1945).
4. Daniel S. Parkinson, School Policy in Student Choices of Courses in High School Mathematics, unpublished Ph. D. Dissertation, University of Wisconsin, 1959.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

TOWARD A THEORY OF INTERPERSONAL RELATIONSHIPS IN THE ADMINISTRATION OF HIGHER EDUCATION

DAVID L. McKENNA
Spring Arbor Junior College
Spring Arbor, Michigan

THE INTERPERSONAL relationship is at the heart of the administrative process. Out of this dynamic interaction between personalities arise the success as well as the failure of administrative leadership. Practicing administrators are quick to observe that the superiority of an institution, the efficiency of its operation and the satisfaction of its faculty are somehow related to the quality of the communication between persons. Yet, no one has objectively defined the difference between administrator-faculty relationships which foster effective communication and those which create tension. Furthermore, the meager research in the area has not provided a theory about the interpersonal relationship which is readily applicable to the everyday problems of the administrator.

This report is an attempt to initiate research in this area by (1) providing the framework for a theory about interpersonal relationships in the administration of higher education; and (2) identifying some characteristics which administrators may use to determine the quality of their relationships with their staff.

I. The Research Design

A. The Problem

Because of the paucity of scientific research and the absence of a theoretical construct for studying interpersonal relationships in the administration of higher education, a field study of an exploratory design was proposed. From among the many administrator-staff relationships, the interpersonal relationship between the departmental chairman and the teaching professor was chosen for the investigation because of the availability of the participants and their key position in translating policy into function. Also, the concept of social power was selected as a segment of the relationship which could be measured and analyzed. Power is defined as "the potential ability of the chairman to influence the behavior of the professor within the area of the chairman's administrative functions." This kind of power is informally delegated to the chairman by the faculty and may or may not be equal to the authority which is institutionally prescribed to the administrative role. The specific problem was to determine if power is a meaningful concept for discussing the interpersonal relationship between the departmental chairman and the teaching professor in higher education.

B. The Procedure

The population for the study was chosen from among the departmental chairmen (or department heads) and teaching professors in four public institutions of higher education. Two of the institutions had enrollments between 15 and 20,000 students. The other two had enrollments under 6,000. A questionnaire survey was used to obtain the data. The final sample was drawn from 47 chairmen and 115 professor respondents.

In his questionnaire, the chairman was asked to check the amount of power that he thought he had for each of his administrative functions on a nine-point scale which ranged from "No Power" to "Complete Power". The functions included were: (1) planning; (2) organizing; (3) staffing; (4) directing; (5) coordinating; (6) reporting; and (7) budgeting. After completing this section, he was asked to indicate the amount of power that he preferred to have for each of these functions. A similar questionnaire was given to the professor so that he checked the amounts of power that he thought the chairman had and the amounts that he preferred him to have. In addition, the chairman was asked to complete a section of forced-choice items which measured whether his style of administrative leadership tended to be personal or impersonal. A personal style stressed the importance of leadership through interpersonal and group relationships; and impersonal style was characterized by reliance upon institutional policy and the administrator's self-sufficiency.

C. The Analysis

From the data which were collected, coded and scored, the analysis was guided by three research questions from which implications about interpersonal relationships could be made. These questions and the analytical procedures which were used for the investigation include:

1. What is the administrative role of the chairman in terms of the amounts of power he is perceived to have for his functions? To analyze the role of the chairman in terms of power, mean scores were obtained for the amounts of power that the chairmen perceived themselves to have for their functions. The functions were then ranked according to their mean scores and an analysis of variance test was conducted to determine whether or not the means were significantly different. An identical procedure was used with the professor's perceptions of the chairman's power.

In order to compare the amounts of power that the chairmen and professors perceived for the functions, T-tests were computed between the mean scores for each function. The tests showed that there were no significant differences between the amounts of power that the chairmen and professors perceived the chairman to have for his functions. Consequently, a mean of the two perceptions was computed for each function and used as an index of the chairman's administrative role in terms of power. Then, the functions were labelled as "High," "Medium" or "Low" power functions according to the mean amount of power ascribed to them.

2. What is the relationship between the amounts of power that the chairman perceives and prefers himself to have for his functions and his style of administrative leadership? Chi-square tests were used to determine whether or not there was a significant difference between the leadership style of the chairman who perceived high power for a function and the one who perceived low power. The same procedure was used to analyze the chairman's preferences for power and his leadership style.

3. What is the relationship between the chairman and professor's perceptions of power and their preferences for power? A procedure similar to the one used in the first area of investigation was carried out with the chairmen and professor's preferences for power in order to obtain mean scores. Then, the chairmen and professor's perceptions of power were compared with their preferences by T-tests. (Because the respondents checked their preferences for power on a scale where their perceptions of power were visible, some artifactual relationship between the variables undoubtedly exists and limits the results.)

Chi-square procedures were used to determine if there were any significant relationships between: (1) the chairman's preferences for power and the size of the institution with which he was affiliated; and (2) the professors' preferences for power and the amount of power the chairman has in contrast with the index power.

D. Research Findings

1. The Administrative Role of the Department Chairman.

a. The chairmen and professors perceived that the chairman has varying amounts of power for his administrative functions which are significantly different (analysis of variance--.05).

b. T-tests indicated that there are no significant differences between the amounts of power that the chairmen and professors perceive the chairman to have for his functions. On the basis of this finding, an index of the chairman's power was constructed with the following descriptions

Reporting--High Power
Budgeting--High Power
Organizing--Medium Power
Coordinating--Medium Power
Staffing--Medium Power
Directing--Medium Power
Planning--Low Power

2. The Chairman's Power and Style of Administrative Leadership.

a. Of the seven administrative functions studied, only the amount of power that the chairman perceived himself to have for the planning function was significantly related (chi-square--.05) to his style of administrative leadership. The chairman who perceived that he had low power for planning tended to be more concerned with personal and group relationships than did the chairman who perceived that he had high power for planning. This finding was further supported by an analysis of the chairman's perception of power and the individual areas which make up leadership style. The chairman who perceives that he has high power for planning tends to be more impersonal toward the faculty member than does the chairman who perceives that he had low power (chi-square--.05). Also, there was some suggestion that the chairman who perceives that he has high power for planning was more personal in relationships with the higher administration than the chairman who perceives low power. However, this finding was significant at only fifteen percent level of confidence.

b. No significant relationships were found between the chairman's preferences for power and his style of administrative leadership.

3. The Chairman and Professors' Preferences for Power.

a. The preferences for power of the chairmen in the large institutions differed significantly (chi-square--.05) from the preferences of the chairmen in the smaller institutions. While the former chairmen preferred less power for the functions of planning, organizing and directing, the chairmen in the smaller schools preferred more power for these functions.

b. The professors' preferences for the chairman's power were significantly related (chi-square --.05) to the amount of power that they perceived him to have in comparison to the index of power. If they perceived that the chairman had much more power for a function than was ascribed in the index of power, the professors preferred that the chairman have less power. Conversely, when they perceived that the chairman had much less power than was provided in the index, they preferred that the chairman have more power for that function.

4. The Interpretation of the Findings.

From the previous findings, inferences were made to provide some practical applications of the study to the administration of higher education and to establish some theoretical guideposts for further research in this area.

II. The Administrative Role of the Departmental Chairman

The departmental chairman is an administrator in higher education whose role is institutionally prescribed to include the functions of: (a) planning; (b) organizing; (c) staffing; (d) directing; (e) coordinating; (f) reporting; and (g) budgeting. These administrative functions require varying degrees of interaction between the chairman and professors and involve decisions for which the chairman must have some power to influence the behavior of the professor. The amount of power that the chairman has depends upon the nature of the function being performed and whether or not the function is expected to be shared with the professors. Therefore, in the description of chairman's role, the high power functions of reporting statistical information and compiling budgetary data are functions which are clerical in nature and are not expected to be shared with the professors. At the other end of the index is the function of planning departmental policies and programs for which the chairman has low power to influence the behavior of departmental staff member. Planning is the goal-setting function for the department and is expected to be shared with those whose welfare is affected by the plans. The medium power functions of organizing, coordinating, staffing and directing are post-planning functions which implement planning in administrative action.

III. The Chairman's Power and His Style of Administrative Leadership

The low power function of planning is the critical function in the administrative role of the chairman for understanding his style of leadership and for discussing the chairman-professor interpersonal relationship. The chairman who perceives that he has high power to influence the behavior of the professor in the area of planning tends to be associated with a more impersonal style of leadership. That is, he: (1) avoids interpersonal relationships with the faculty; (2) protects his own status and position by discouraging the professional growth of the professor; (3) finds the source authority for his leadership in the higher administration; and (4) prefers a type of chairman who is known for his acquaintance with institutional policy and his practical ability to get things done. On the other hand, the chairman who shares the planning function with the professors, (1) fosters interpersonal relationships with the faculty; and (2) encourages the professional growth of the professors; (3) finds the source of authority for his leadership in the departmental faculty, and (4) prefers the type of chairman who is known for his understanding of the professor's problems and for his reputation as a leader of the faculty.

IV. The Chairman and Professor's Preferences for Power

The chairman's preferences for power for his functions tend to be institutionally-related, while the professor's preferences are more personal. Whether or not the chairman prefers more or less power for his functions depends upon the size of the institution with which he is affiliated. A departmental chairman in the larger institution has become the administrator of a relatively autonomous unit within the organization. Consequently, he has more power and responsibility than he prefers to have for planning, organizing and directing because he does not consider himself a professional administrator. On the other hand, the chairman in the rapidly-expanding institution has been given additional administrative responsibilities, but he does not yet have the power to carry out these functions effectively. Having moved so rapidly from a clerical to a full-fledged administrative position, the chairman in the smaller institution prefers more power for the important functions of planning, organizing and directing.

In contrast, the preferences of the professors are directed toward a balance of power in which the chairman performs his functions, but does not usurp the prerogatives of the faculty. When the professor perceives that the chairman has much more power than is sanctioned by the index of power, he prefers that the chairman's power be reduced. Conversely, when the chairman is perceived to have much less power than prescribed by the index, the professor prefers that he have more power. By his preferences, professor seeks to restrain the high power chairman from overexercising his power and puts pressure on the low power chairman to assume more power.

This is the basis for suggesting that the preferences for power of the chairman are institutional while the preferences of the professor are personal. The chairman, whose position of power lends some personal security, is primarily interested in maintaining the power that will enable him to be efficient in his institutional role. The professor, on the other hand, is concerned with achieving or retaining his personal security and, therefore, seeks to guard against the imperious threat of the autocrat or the vacillation of the indecisive chairman.

V. The Characteristics of High and Low Tension Levels in the Chairman-Professor Interpersonal Relationship

With the background of the previous findings, certain implications may be drawn about the tension levels within a department which have a direct bearing upon the quality of the interpersonal relationships between chairman and professors.

A low tension level exists within a department when:

1. The chairman shares the low power function of planning departmental programs, policies and regulations with the professor.
2. The chairman recognizes the affective needs of the faculty and shares goal-setting functions of the department with the professors when their personal welfare is concerned.
3. The chairman finds the source of authority for his leadership in the departmental faculty.
4. The chairman is secure in his administrative position and encourages the professional growth of the professor.
5. The discrepancy between the chairman's power and the index of power is low.

A high tension level exists within a department when:

1. The chairman does not share the low power function of planning departmental programs, policies and regulations with the professors.
2. The chairman is more concerned about maintaining his own position and the efficiency of his department than he is about the affective needs of the professors; he does not share the goal-setting functions of the department with the professors even when their own personal welfare is involved.
3. The chairman finds the source of authority for his leadership in the higher administration.
4. The chairman is insecure in his administrative position and discourages the professional growth of the professor.
5. The discrepancy between the chairman's power and the index of power is high.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

SELECTED OUT-OF-SCHOOL FACTORS THAT AFFECT NEGRO ELEMENTARY SCHOOL CHILDREN

ELSIE HILL WALLACE
Florida A. and M. University

THE MODERN elementary school exists in a cultural-social situation where mass media, the most powerful teachers of our generation, are demanding more and more of the child's time. The elementary school, therefore, is in keen competition with these out-of-school factors, and they are as potent, if not more compelling, than the in-school experiences.

Parents and teachers of elementary school children are faced with the responsibility of guiding their young so that they will make wise choices, in leisure time pursuits, as well as further develop their capacity for judgment, selection and appraisal of cultural offerings.

Purpose of the Study

The purpose of this study was to ascertain the percent of Negro elementary school children in Leon County, Florida, who (a) observed television programs, (b) listened to radio programs, (c) read comic books, (d) attended movies, (e) participated in club activities, and (f) attended Sunday school during the year 1958-59.

Method

The responses of three thousand six hundred and eighty-two (3,682) Negro elementary school children in Leon County, Florida, were tabulated, compiled, summarized and evaluated. The children were enrolled in ten elementary schools, comprising grades one through six.

Of the 3,682 Negro elementary school children enrolled in the elementary schools of Leon County, Florida, 60.2 percent observed television programs and 76.8 percent listened to radio programs during the year 1958-59. The highest percent of television observation was in the urban areas while the highest percent of radio listening was in the rural areas.

A classification of the programs showed that the children listened to more "Action Drama," "Music," and "Human Interest" programs than to "Humor" and the "News." The programs were classified as quoted above.

Data in Table II show that 52.3 percent of the Negro elementary school children attended the movies and 57.9 percent read the comics. The highest percent of movie attendance was in the urban areas and the rural schools located nearest to the metropolitan centers.

While the majority of comics were classified under "Action Drama," the general choice indicated a variety of interest. The "Classics" and "Human Interest" comics were preferred over the "Detective" or "Adventure." Animal stories ranked high among the choices, especially in the primary grades.

Data in Table III indicate that only 21.9 percent of the Negro elementary school children participated in club activities while 97.3 percent attended Sunday school during the year 1958-59. The highest percentage of club participation was in the Southeast School (an urban school) and Barrow Hill School (a rural school) with the lowest percent of participation in the Griffin School (an urban school) and the Concord School (a rural school).

The percent of children who attended Sunday school was fairly high in each of the schools with the exception of the Lucy Moten School, a laboratory situation located on the University campus.

Findings

The television, radio, comics, movies, club activities and Sunday school were significant out-of-school factors in the lives of Negro elementary school children in Leon County, Florida, during the year 1958-59.

1. Sixty and two tenths percent of the children observed television. The favorite television programs were classified under "Action Drama" with emphasis on the "westerns" and "cartoons."
2. A high percentage of the children listened to radio programs in the rural and urban sections (76.85). The programs mentioned were classified as 1) humor, 2) action drama, 3) human interest, 4) music, 5) the news and 6) melodrama.
3. The data also revealed that 52.3 percent of

TABLE I

PERCENT OF NEGRO ELEMENTARY SCHOOL CHILDREN WHO OBSERVED TELEVISION AND LISTENED TO THE RADIO IN LEON COUNTY, FLORIDA, 1958-59

Type of School*	Name of School	Total School Enrollment	Percent of Children Who Observed Television Programs	Percent of Children Who Listened to Radio Programs
R	Barrow Hill	241	19.9	99.2
U	Bond	691	60.2	40.4
R	Concord	236	16.5	70.3
U	Griffin	936	88.6	76.6
U	Lucy Moten	188	86.2	87.2
R	McBride	240	31.3	92.5
R	Raney	146	30.1	76.0
U	Riley	416	68.8	99.3
U	Southeast	417	62.1	88.0
R	Station One	171	35.1	87.1
Totals		3,682	60.2 (Average)	76.8 (Average)

* R--Rural, U--Urban

TABLE II

PERCENT OF NEGRO ELEMENTARY SCHOOL CHILDREN WHO ATTENDED THE MOVIES AND READ THE COMICS IN LEON COUNTY, FLORIDA, 1958-59

Type of School*	Name of School	Total School Enrollment	Percent of Children Who Attended Movies	Percent of Children Who Read the Comics
R	Barrow Hill	241	67.2	14.5
U	Bond	691	57.9	63.5
R	Concord	236	19.9	16.9
U	Griffin	936	52.1	51.4
U	Lucy Moten	188	82.4	80.8
R	McBride	240	12.1	68.8
R	Raney	146	35.2	55.5
U	Riley	416	58.2	67.8
U	Southeast	417	59.2	74.3
R	Station One	171	20.5	40.8
Totals		3,682	52.3 (Average)	57.9 (Average)

* R--Rural, U--Urban

TABLE III

PERCENT OF NEGRO ELEMENTARY SCHOOL CHILDREN WHO PARTICIPATED IN CLUB ACTIVITIES AND ATTENDED SUNDAY SCHOOL IN LEON COUNTY, FLORIDA, 1958-59

Type of School*	Name of School	Total School Enrollment	Percent of Children Who Participated in Club Activities	Percent of Children Who Attended Sunday School
R	Barrow Hill	241	51.5	83.8
U	Bond	691	25.2	77.7
R	Concord	236	1.2	57.2
U	Griffin	936	3.7	61.6
U	Lucy Moten	188	20.7	38.8
R	McBride	240	37.9	90.0
R	Raney	146	25.7	51.4
U	Riley	416	25.7	89.7
U	Southeast	417	52.5	95.7
R	Station One	171	0.0	56.7
Totals		3,682	21.9 (Average)	97.3 (Average)

* R--Rural, U--Urban

the children attended the movies. The highest percentage of attendance was in the urban areas and the rural schools located nearest to the metropolitan centers. The favorite movies were classified mainly under "action drama," and "adventure."

4. According to the study, only 21.9 percent of the children participated in club activities. On the other hand, however, 97.3 percent attended Sunday school.

Conclusions

Since the television, radio, comic books, movies and Sunday school are significant out-of-school factors in the lives of Negro elementary school children in Leon County, Florida, parents, teachers and community laymen need to:

1. Study these forces in an effort to determine how they may be utilized more effectively in the learning of children.
2. Help children develop skills of critical thinking so that they can be more discriminating in their choice of programs and activities.
3. Work together cooperatively in planning television and radio programs that are more in keeping with acceptable standards for children's programs.
4. Seek to captivate children's natural interest for reading the comics in developing a taste for literature on a higher level.
5. Encourage more extensive participation in club activities so that the lives of their children may be broadened and enriched to a fuller extent.

REFERENCES

Frank, Lawrence K., "Our Children's In-School and Out-of-School Teachers," Educational Leadership, Washington: N. E. A., Association of Supervision and Curriculum Development, February, 1955, pp. 293-300.

Garry, Ralph, et al., "Television for Children," Journal of Education, Boston: School of Education, Vol. 140, 1957, p. 7.

Herrick and Jacobs, Children and the Language Arts, Englewood Cliffs, New Jersey: Prentice Hall Co., Inc., 1955, p. 378.

Tooze, Ruth, Your Children Want to Read, Englewood Cliffs, New Jersey: Prentice Hall Co., Inc., 1957, p. 28.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

MISSING DATA IN NON-LINEAR TREND ANALYSIS OF REPEATED MEASUREMENTS ON THE SAME INDIVIDUALS*

RICHARD B. SEYMOUR**
Long Beach, California

Problem

THIS PAPER deals with the problem arising in the analysis of trends based on longitudinal data, when not all cases are available in successive measurements. Formulas are presented for estimating missing data in Grant's design (3) for analysis of non-linear trend, making it possible to use the incomplete individuals and apply the same straightforward computations as are used for a complete set of data.

Repeated measurements on the same individuals are made in a number of research areas, such as learning, mental testing, medical therapy and personality. Alexander (1) developed a method of analysis of variance in which deviations from linearity, both group and overall in the case of two or more groups, are tested for significance. Group deviation from linearity is more specifically defined as the overall trial deviations from linearity superimposed upon each of the group slopes or linear trends. Grant (3) extended the method a decade later by partitioning degrees of freedom so as to test the following sources of variation: between group linear trends, between group quadratic trends, between group cubic trends, and so on; and between corresponding individual trends. The degree of polynomial is limited by the number of trials, so that three trials permit testing linear and quadratic trends, for example, and four trials permit testing cubic trends in addition.

In a number of experiments, however, the design is marred by missing data due to accidents, illness and the like. One method of dealing with this situation is to write out the normal equations for the observations that are present and to solve them, but the difficulty and complexity of solution are increased and many competent investigators lack the background for such analysis (2). More widely usable is the method of insertion, in which values are inserted for the missing data and the usual computational procedures followed, then degrees of freedom are adjusted (2). Kendall (7) has presented a worked example with nine missing data in a p × q classification with one observation per cell (except for missing data). Kempthorne (6) gives the computational method for a 2-way table and the basis of bias in F, and suggests use of the biased F test to see whether it would be desirable to calculate the exact value of F. Values may also be guessed and an iterative method followed.

It is much quicker, however, to use a missing data formula and compute as for complete data, then make a minor adjustment in degrees of freedom (method of insertion). The advantage of time-saving is often important, and the facilities of electronic computers for more complex solutions are not always available. The present paper gives formulas for the more common cases arising in practice with Grant's method and discusses their use.

Formulas

The formulas in this paper apply to cases involving one observation per individual per trial and classification of an individual (or plant or object) in one and only one group. The groups need not be of equal size.

Grant (3) uses several error terms, depending on which hypothesis is to be tested, and theoretically any of these error terms might be minimized as a basis for supplying an estimate of missing data. It is immediately apparent, however, that the estimate which minimizes between individual linear trends will not necessarily minimize between individual quadratic trends or higher order trends. For this and other, more complicated reasons, the formulas derived on this basis are not satisfactory.

Since between group trends is frequently the basic source of variation to be tested for significance, it sould seem reasonable to minimize between individual trends, the error term for testing between group trends, rather than to minimize a specific component of trend. In addition, between individual trends has more degrees of freedom associated with it and hence is less affected by one or more aberrant values.

If one datum is missing, or if two or more are missing but each is from a different group (while the trials may be all the same or not all the same), the missing datum is estimated by formula (1) based on minimization of between individual trends:[1]

---

* Footnotes will be found at the end of the article.

$$X_1 = \frac{n_1 I_1 + kT_{11} - G_1}{(k - 1)(n_1 - 1)} \tag{1}$$

where

n_1 = number of individuals in the group
k = number of trials
I_1 = incomplete sum of scores for individual with missing datum
T_{11} = Incomplete sum or scores for earlier trial with missing datum in group with missing datum
G_1 = incomplete sum of scores for group with missing datum

When the two missing data are from the same individual (but by the nature of the restrictions stated above, are at two different trials), formulas (2) and (3) estimate the missing data.

$$X_1 = \frac{(k - 1) \quad (n_1 I_1 + kT_{11} - G_1) + n_1 I_1 + kT_{12} - G_1}{(n_1 - 1)\, k\, (k - 2)} \tag{2}$$

where

I_{11} = incomplete sum of scores for earlier trial with missing datum in group with missing datum, e. g., trial 1 and group 1
T_{12} = incomplete sum of scores for later trial with missing datum in group with missing datum, e. g., trial 2 and group 1

$$X_2 = \frac{n_1 I_1 + kT_{12} - G_1 + (n_1 - 1) X_1}{(k - 1)(n_1 - 1)} \tag{3}$$

Although a formula has been derived for X_2 which is independent of X_1, formula (3) is considerably more convenient.

When the two missing data are from two different individuals, in the same group at the same trial, formulas (4) and (5) are appropriate:

$$X_1 = \frac{(n_1 - 1)(n_1 I_1 + kT_{11} - G_1) + n_1 I_2 + kT_{11} - G_1}{(k - 1)(n_1^2 - 2n_1)} \tag{4}$$

where

I_2 = incomplete sum of scores for second individual with missing datum

$$X_2 = \frac{n_1 I_2 + kT_{11} - G_1 + (k - 1) X_1}{(k - 1)(n_1 - 1)} \tag{5}$$

For the case of two individuals in the same group but at different trials, the following two formulas apply:

$$X_1 = \frac{1}{[(k - 1)(n_1 - 1)]^2 - 1}\left[(k - 1)(n_1 - 1)(n_1 I_1 + kT_{11} - G_1) - n_1 I_2 - kT_{12} + G_1\right] \tag{6}$$

$$X_2 = \frac{n_1 I_2 + kT_{12} - G_1 - X_1}{(k - 1)(n_1 - 1)} \tag{7}$$

The loss of a third observation considerably complicates the formulas, with the exception of three cases: each observation is from a different group (and therefore also a different individual), while the trials may be 1) all the same, 2) two the same, or 3) none the same. The formula for these three cases is formula (1), with appropriate modifications being made in the subscripts identifying group, trial or individual.

Once the estimates are made, the analysis of variance is carried out with the computational methods which are used for complete sets of data, as with the method of insertion in other designs of analysis of variance. Grant (3) gives both the computational steps and a worked example of complete data. The only modification in most cases is that one degree of freedom must be deducted from the error term, between individual trends, for each estimated observation. Consequently somewhat higher values of F must be attained for significance at the chosen level.

The use of the formulas is illustrated in Tables I and II. Table I presents unpublished data of Harold E. Jones on electrical skin resistance of adolescents measured at the beginning of experimental sessions in the California Adolescent Growth Study. The logarithmic transformation of ohms resistance was used in order to meet the assumptions of analysis of variance. Further details are given by Jones (4, 5).

Discussion

The usual adjustment of degrees of freedom discussed above is adequate for most cases and is employed in other analysis of variance designs. Clearly significant F tests and those based on large numbers of degrees of freedom are essentially unaffected by the adjustment. In tests of borderline significance based on few degrees of freedom, however, the adjustment may be less satisfactory. In such cases, the possibility of bias in the direction of more significant F tests should be borne in mind. While the bias is usually considered small, Scott (8) and Snedecor (10) have found in similar situations with other designs that it is sometimes considerable. In cases of doubt, more cases may be added while retaining the incomplete individuals, or the incomplete individuals may be discarded.

The generality of the formulas for cases of more than one missing datum remains to be specified. Subscripts may be interchanged as long as the verbal description fits. Thus, if two missing data are from two individuals in one group at two trials, the group, trial and individual (ikm) may be designated

TABLE I

ELECTRICAL SKIN RESISTANCE IN LOG OHMS AT SIX-MONTHLY INTERVALS

Case Number	Trial Number						Total
	1	2	3	4	5	6	
104	5.256	5.157	4.977	5.324	5.301	5.276	31.291
114	5.253	5.138	5.296	5.073	5.210	5.127	31.097
118	5.172	5.086	5.123	4.757	4.710	X_1	(24.848)
172	5.047	5.074	4.809	5.084	4.913	4.367	29.294
Total	20.728	20.455	20.205	20.238	20.134	(14.770)	(116.530)

Note: Grateful asknowledgement for permission to use the data is made to Dr. Harold E. Jones, Director of the Institute of Child Welfare, University of California, Berkeley. The data are only a small fraction of those gathered and do not necessarily represent the true trends.

TABLE II

ILLUSTRATION OF USE OF FORMULAS (1)-(7) WITH DATA OF TABLE I

Formula	Case	Trial	Designations	Estimate
1	118	6	X_1 or X_{163}	4.7655
2	118	4	X_1 or X_{143}	5.0638
3	118	6	X_2 or X_{163}	4.8268
4	114	6	X_1 or X_{162}	4.9213
5	118	6	X_2 or X_{163}	4.6961
6	114	3	X_1 or X_{132}	5.1097
7	118	6	X_2 or X_{163}	4.8335

Note: While only one observation is actually missing from Table I, the cases of two missing data are illustrated with the same table by assuming another observation to be missing, as the need arises. Designations are given both as first (X_1) or second (X_2) missing datum and as X_{ikm}, where i = group, k = trial, and m = individual, and individuals are numbered consecutively, case 104 being individual 1.

111 and 222 or 121 and 112, using only 1 and 2 as the simplest designations for trials and individuals, though it might equally well be 5 and 8 or any other pair of designations.

Similarly, for the three cases of three missing data to which formula (1) is applicable, designations may be 111, 222 and 333, or 112, 223 and 331, or 122, 233 and 311, and so on, but not 122, 133 and 312, because in the last instance the first two missing observations are from the same group, namely 1, and because the first and third observations are from the same individual, namely 2.

The latter error not only disqualifies the three data from being suitable for formula (1) but also is impossible in this design since a given individual can appear in only one group. Designations 213, 312 and 111 represent data which are suitable for formula (1) since it is of no consequence whether the data are from the same trial or not, when there are three individuals in three groups. In the case of two missing data, there are eight theoretically possible cases of sameness-differentness of group, individual and trial, but only six can exist due to restrictions of design. Similarly, for three missing data there are 26 theoretical possibilities but only 10 can exist.

Additional cases for which no formula is presented above arise occasionally, such as three missing data in the same group. Resources did not permit the inclusion of these additional cases.

Summary

Formulas for missing data to permit use of conventional computational methods are derived and discussed for analysis of repeated measurements on the same individuals (including objects). The formulas are for use with the orthogonal polynomial analysis of trends, an extension of Alexander's method due to Grant. Formulas are presented for the case of one missing datum in Grant's and Alexander's methods and for all cases of two missing data in Grant's method, minimizing the error term, between individual trends, in the latter method. The formulas are applicable by extension and without modification to certain cases where more than two observations are missing. An example is presented.

FOOTNOTES

* Grateful acknowledgement is made to Mr. Atholl L. Wilson, Department of Statistics, University of California, Berkeley, and to Professor G. W. Snedecor, Statistical Laboratory, Iowa State College, for their helpful suggestions and criticisms.

** Now at 1124 West Carson St., Torrance, California.

1. The formulas were derived by minimizing between individual trends with the use of calculus. It is interesting to compare this formula with the formula for one missing datum in Alexander's method, based on minimizing individual deviations from estimation, also by calculus:

$$X_1 = \frac{n_1 I_1 \Sigma x^2 + kT_{11}\Sigma x^2 - kP_{1\cdot} + kn_1 P_{11} - G_1 \Sigma x^2}{(n_1 - 1)\left[k(\Sigma x^2 - x) - \Sigma x^2\right]}$$

where x = time multiplier associated with the missing datum, Σx^2 = sum of squares of time multipliers, P_{11} = incomplete sum of scores, each times its time multiplier (for individual with missing datum), $P_{1\cdot}$ = incomplete sum of scores, each times its time multiplier, for group with missing datum.

REFERENCES

1. Alexander, H. W., "A General Test for Trend," Psychological Bulletin, 1946, 43, 533-557.
2. Cochran, W. G. and Cox, Gertrude M., Experimental Designs, New York: Wiley, 1950.
3. Grant, D. A., "Analysis-of-Variance Tests in the Analysis and Comparison of Curves," Psychological Bulletin, 1956, 53, 141-154.
4. Jones, H. E., Development in Adolescence, New York: D. Appleton-Century, 1943.
5. Jones, H. E., "The Study of Patterns of Emotional Expression," in M. L. Reymert (Ed.), Feelings and Emotions. New York: McGraw-Hill, 1950.
6. Kempthorne, O., The Design and Analysis of Experiments. New York: Wiley, 1952.
7. Kendall, M. G., The Advance Theory of Statistics. London, Griffin, 1946. Vol. 2., pp. 229-233.
8. Scott, Elizabeth L., Personal Communication, 1957.
9. Snedecor, G. W., Statistical Methods. (5th Ed.) Ames, Iowa: Iowa State College Press, 1956, pp. 312-313.
10. Snedecor, G. W., Personal Communication 1957.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

SPECIAL BUDGETARY PROVISIONS MADE FOR EDUCATING THE GIFTED

MORRIS J. APPLBAUM
Brooklyn, New York

IN A RECENT SURVEY* pertaining to the education of 1956 Merit Certificate Winners, questionnaires sent to principals of schools attended by these winners revealed that in most instances no special budgetary provisions were made for educating the gifted. Prominent among the questions asked was: "Are special budgetary provisions made for use in the education of the gifted?" The principals were requested to rate the practice of making special budgetary provisions for educating the gifted in three ways: first, in terms of how often or how persistently the practice was utilized in the high school; second, in terms of how successful the practice was found to be when utilized; third, in terms of how valuable the practice would be without regard for the use which was being made of the practice.

Tables I, II, and III show a detailed breakdown of the responses of the 1,063 principals who returned their questionnaires. Table I provides the data when the responses regarding the evaluation of the practice were tabulated separately; Table II provides the data when the responses on utilization and evaluation of the practice were combined for tabulation purposes; and Table III provides the responses to the theoretical evaluation.

Table I shows: that 78.3 percent of the principals indicated that the practice of providing special budgetary allowances for use in educating the gifted was utilized very little or not at all; that 10.2 percent indicated that it was utilized to a moderate extent; and that 2.8 percent indicated that it was utilized to a considerable extent. Furthermore, from Table I can be gathered that 4.2 percent principals found the practice very successful, 13.3 percent moderately successful, and 54.4 percent unsuccessful. Table II, however, provides a more revealing picture. Here 2.2 percent of the respondents indicated that the practice was utilized to a considerable extent with great success and .6 of one percent to a considerable extent with moderate success. Conversely, 1.2 percent reported that the practice was utilized to a moderate extent with great success and 8.4 percent to a moderate extent with moderate success.

Table III actually shows what the principals think of providing special budgetary allowances for use in teaching the gifted. Although this constitutes a theoretical evaluation of the practice, it is of great significance when compared with the findings in Tables I and II. Here, in Table III, 36.1 percent of the principals responding feel that special budgetary provisions would be of considerable value; 30.5 percent feel that they would be of moderate value; and 18.7 percent feel that they would be of little or no value.

It is interesting to observe that although 36.1 percent of the principles feel that special budgetary provisions would be of considerable value, only 2.8 percent made such provisions in their schools.

---

* Morris J. Applbaum, *A Survey of Special Provisions for the Education of Academically Superior Students in the United States* (Unpublished Doctoral Dissertation, Teachers College, Columbia University) 1958, 388 pp. See *Phi Delta Kappan*, Vol. XXXX, No. 8, May, 1959.

TABLE I

RESPONSES FROM PRINCIPALS REGARDING UTILIZATION AND EVALUATION OF PRACTICE OF MAKING SPECIAL BUDGETARY PROVISIONS FOR USE IN THE EDUCATION OF THE GIFTED

Respondents (Principals)	Responses Regarding Utilization of Practice				
	Used Considerably	Used Moderately	Used Little or Not at All	No Response	Total
Number of	30	108	832	93	1,063
Percent of	2.8	10.2	78.3	8.7	100
	Responses Regarding Evaluation of Practice				
	Very successful	Moderately Successful	Unsuccessful	No Response	Total
Number of	45	141	578	299	1,063
Percent of	4.2	13.3	54.4	28.1	100

TABLE II

COMBINED RESPONSES OF PRINCIPALS ON UTILIZATION AND EVALUATION OF PRACTICE OF MAKING BUDGETARY PROVISIONS FOR USE IN THE EDUCATION OF THE GIFTED

Respondents (Principals)	Considerably Used, Very Successful	Considerably Used, Moderately Successful	Much Used, but not Very Successful	Moderately Used, Very Successful	Moderately Used, Moderately Successful	Moderately Used, not Very Successful	Little Used, Very Successful	Little Used, Moderately Successful	Little Used, not Very Successful	Not Used, so Not Evaluated	No Response	Total
Number of	23	7	--	13	89	6	9	45	572	206	93	1,063
Percent of	2.2	0.6	--	1.2	8.4	0.6	0.9	4.2	53.8	19.4	8.8	100

TABLE III

RESPONSES FROM PRINCIPALS ON THEORETICAL EVALUATION OF PRACTICE OF MAKING BUDGETARY PROVISIONS FOR USE IN THE EDUCATION OF THE GIFTED

Respondents (Principals)	Of Considerable Value	Of Moderate Value	Of Little or no Value	No Response	Total
Number of	384	324	199	156	1,063
Percent of	36.1	30.5	18.7	14.7	100

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

AN ATTENTION SCALE FOR EVALUATING E.T.V. PROGRAMS

CECIL C. BRIDGES, JR.
Oklahoma City Public Schools
Oklahoma City, Oklahoma

IN EVALUATING film or television teaching, a measure of pupil attention is sometimes useful. In the Oklahoma City E. T. V. experiment, for example, there was some disagreement about the appropriate length for E. T. V. programs for elementary pupils. It was felt by some that the optimum program length might be shorter than thirty minutes. In some classes which had been observed, pupil attention seemed to drop markedly after about twenty minutes.

Procedure

In order to evaluate pupil attention at intervals over a thirty minute period, a procedure similar to the T. V. and radio rating poll was used - the proportion of the pupils in a class who were watching at specified times was used as an index of pupil attention.

In preliminary observations, it was soon discovered that it was easier to count the pupils not watching than to count those watching an E. T. V. program. The following graphs were calculated from counts made by teachers at five minute intervals while their classes were viewing E. T. V. programs. All of the data were collected in three large elementary schools participating in the Oklahoma City E. T. V. experiment.

Class size ranged from twenty-eight to one hundred pupils; most of the pupils were in large "special area" viewing classes of about seventy pupils each.

Results

Figure 1 shows the proportion of pupils watching at five minute intervals.

Although there is an overall decline in pupil attention as a function of time, the proportion of pupils not watching has not reached serious proportions by the end of these thirty minute programs. In Figure 2, attention is related to grade level.

Since only three E. T. V. programs were seen by more than one class, program material is not constant for this comparison, and these differences between grades can only be considered suggestive.

The validity of this method is assumed, since "watching" is a behavioral definition of attention. The "reliability" of the method depends on how reliability is defined. In most cases, it is not hard for an observer to decide if a pupil is watching or not. High "reliability" could be demonstrated by a correspondence in counts made by different observers of the same class, but this would be rather trivial. Of more interest is the manner in which the fluctuations of pupil attention are correlated between different classes watching the same program. In Figure 3, three different programs are represented, each viewed by three different classes.

A different observer made the counts in each of these classes. It is apparent that in addition to the variations between individual classes, there is a general pattern which varies from program to program. "See It, Say It, Do It," March 4, is characterized by a slight dip in pupil attention from five to fifteen minutes, followed by an upswing of attention at twenty to twenty-five minutes. "Wonder World," February 11, seemed to hold the attention of all pupils fairly well up to twenty-five minutes, after which there was a sharp decline. In "Wonder World." February 25, a low point was reached at fifteen minutes with attention high at beginning and end.

On the Television Attention Scale blank, space was provided for the observer's remarks under "points at which attention was lost" and "points at which attention was held." This is the kind of "feed back" information which is useful in evaluating the efficiency of teaching techniques in terms of pupil attention. The same kind of information is obtained regularly on a "feed back" sheet in the Oklahoma City E. T. V. experiment, but in the case of the Television Attention Scale, such remarks are associated with a quantitative index of pupil attention.

How consistent is the variation of attention during a program from one class to another? Pearson product-moment correlation coefficients between proportions of students viewing the same program were calculated. Such a correlation over program time measures the tendency for attention levels in

FIGURE 1

PROPORTION OF STUDENTS WATCHING AS A FUNCTION OF PROGRAM TIME
(N = 747)

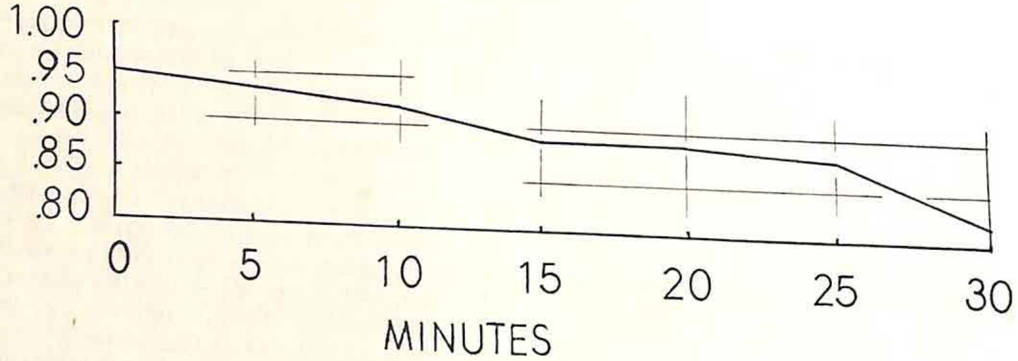


FIGURE 2

PROPORTIONS IN GRADES 4, 5, AND 6 WATCHING AS A FUNCTION OF PROGRAM TIME

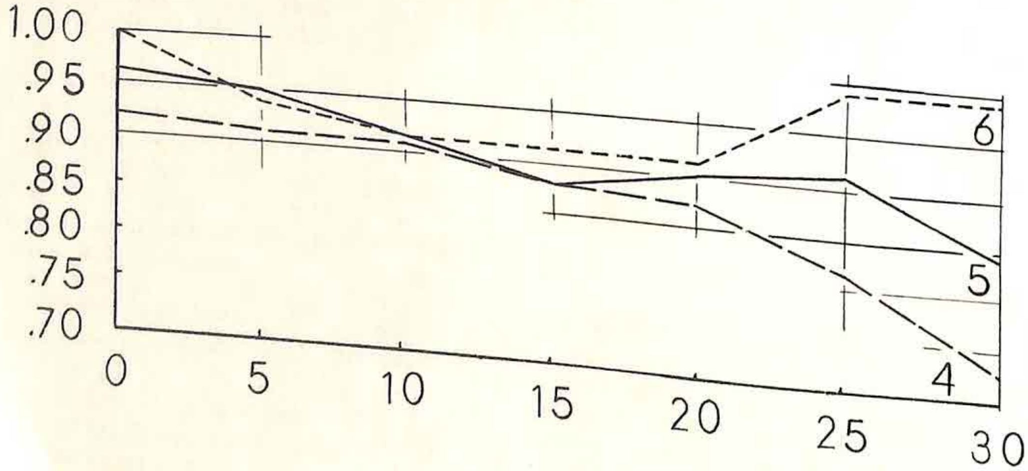
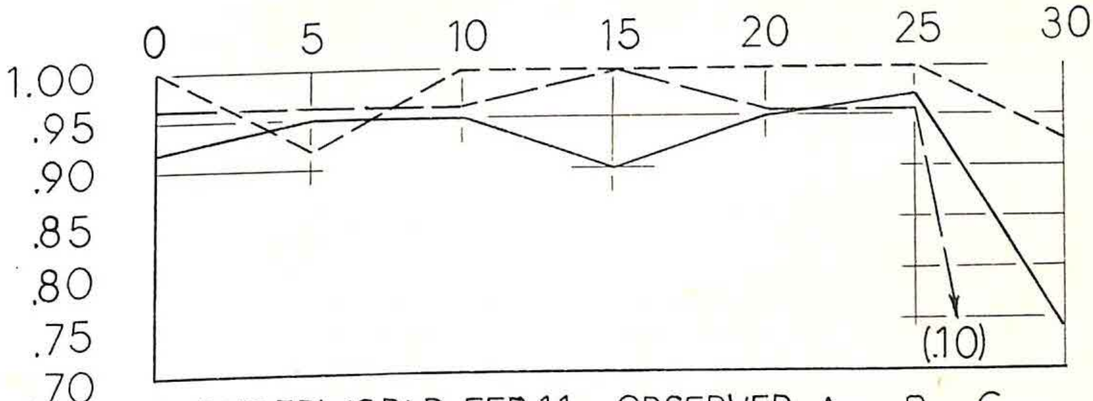
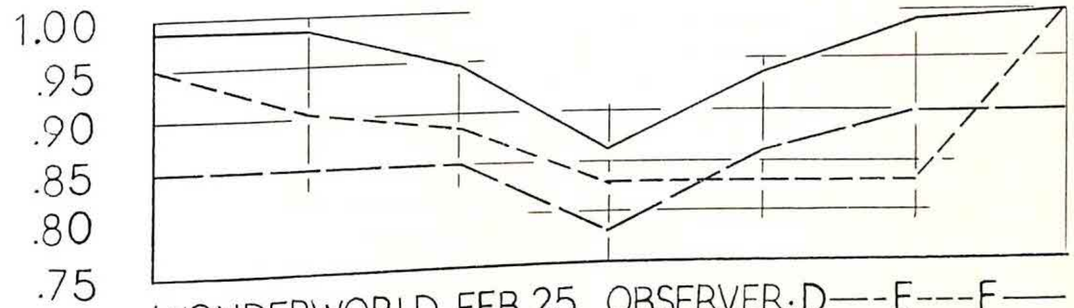


FIGURE 3

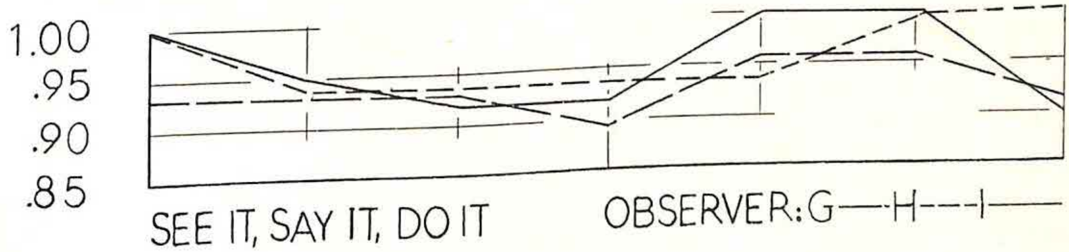
ATTENTION RELATED TO SPECIFIC PROGRAMS



WONDERWORLD, FEB. 11. OBSERVER: A—-B---C—



WONDERWORLD, FEB. 25. OBSERVER: D—-E---F—



SEE IT, SAY IT, DO IT OBSERVER: G—-H---I—

the two classes to vary together. These correlations are as follows:

$r_{ab} = .652$ $r_{de} = .381$ $r_{gh} = .065$

$r_{ac} = .945$ $r_{df} = .862$ $r_{gi} = .784$

$r_{bc} = .574$ $r_{ef} = .654$ $r_{hi} = .190$

A great majority of the pupils were watching most of the time. The similarities between different classes watching the same program indicate that this index of pupil attention depends, in a large degree, on individual program characteristics.

Summary and Conclusions

A Television Attention Scale, consisting essentially of a poll of students watching at five minute intervals, was constructed to evaluate the loss of pupil attention as a function of viewing time. A general decline of pupil attention as a function of viewing time was found, but this decline did not reach serious proportions within thirty minutes. The index of attention (proportion watching) varies consistently during programs within pairs of classes watching the same program. Feed back remarks written on the Television Attention Scale blank with a quantitative index of fluctuating pupils attention are more meaningful than such remarks alone.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

THE EFFECTS OF THE PURPORTED VALIDITY OF A TEST UPON LEVEL OF MOTIVATION AND PERFORMANCE

LAWRENCE S. WRIGHTSMAN, Jr.
George Peabody College for Teachers
Nashville, Tennessee

DESPITE THE concern in educational psychology today over lack of student motivation and the problems of under-achievement, we know little about the level of student motivation in the actual testing situation. A survey of recent texts in educational measurement and evaluation (for example, 3, 5, and 8) and in educational psychology (2 and 9) indicates that the topic is usually ignored, and the general feeling seems to be that on the basis of observations of students' test-taking behavior, we may assume that they are well-motivated. It is of interest, however, to consider what effect purported characteristics of the testing instrument have upon students' motivation and performance.

In one of the more fruitful approaches to this problem, McClelland and his associates (7, Chapter V) have found that in regard to simple tasks the strength of achievement motivation in some subjects may be experimentally aroused and manipulated through the use of instructions which focus on the importance or unimportance of the task. Other studies (1 and 6) indicate that for some simple laboratory tasks the level of performance is related to the strength of the S's typical (i. e., not experimentally aroused) level of achievement motivation.

The purpose of this study is to apply this approach to a more complex task -- the taking of an intelligence test -- and to study the effect of a specific test characteristic -- the validity of the test -- upon the motivation and performance of the students. On the basis of the previous research using simple tasks, it is hypothesized that students who believe the intelligence test to be a valid one will perform better on it than will students who believe it to be invalid.

Method

Forty undergraduates at a midwestern state university were used as subjects (Ss). Ss were enrolled in a laboratory psychology course dealing with tests and measurements. At the time of the experiment, which took place during a regular class period, Ss had been exposed to the theory of measurement, types of validity and reliability, and the characteristics of different tests. About ten tests were taken by the students as part of the course, and the testing done as a part of this experiment was apparently perceived by the Ss as one of the routine classroom procedures.

The Adaptability Test, by Tiffin and Lawshe, was used. A 35-item test designed to measure "mental alertness" (10), it has a usual time limit of 15 minutes.

Each S was given a copy of the Adaptability Test. Instead of recording their answers on the test booklet, Ss were told to use the special mimeographed answer sheet. Ss were requested to put their names on the answer sheet.

Ss in the first and last rows were in the "Valid Test" condition and received an answer sheet on which was printed:

"The Adaptability Test was devised at the Occupational Research Center at Purdue University. The test is an excellent measure of job success for college graduates. The Adaptability Test measures a complex of emotional, motivational, and intellectual factors very important in most occupations. . . . This adaptability factor, involving speed and accuracy in judgment, has been found to correlate .65 with occupational success, using a group of 700 Indiana college graduates as a sample. Thus it is a valid measure of a crucial ability in job success."

Those Ss in the middle two rows were in the "Invalid Test" condition and received an apparently similar answer sheet which really differed from that for the "Valid Test" condition in the following ways:

a) The second sentence read ". . . attempts to be a measure. . ." rather than ". . . is an excellent measure. . ."

b) The third sentence read ". . . the Adaptability Test tries to measure. . ." instead of simply ". . . measures. . ."

c) The validity coefficient with occupational

TABLE I

ANALYSIS OF VARIANCE OF MEAN SCORES ON THE ADAPTABILITY TEST

Source of Variance	df	Mean Square	F
Validity of Test	1	112.2	6.74*
Availability of Answers	1	1.2	.07
Validity x Availability	1	10.3	.62
Error	36	16.65	

*p .05

success was "only .05" rather than .65.

d) The last sentence was changed to read, "Thus it is not considered a valid test for measuring job success."

The experimenter (E), who was the instructor, requested that each student read the information carefully and indicate by looking up when he had completed reading it. E did not discuss the material, and no S was aware of the differences in the purported validity of the test.

Before the test booklets were distributed, half were filled in with the correct answers and then the answers were erased, though deliberately left visible. These booklets were alternated with unmarked ones, so that on each row alternate students did and did not have the answers available. As the items began with easy ones and became progressively harder, it was apparent at least for the beginning items that the half-visible answers were the correct ones. The purpose of this contrivance was to create a situation for half the subjects in which motivation level would be more directly related to performance. For if there is a relationship between purported validity of the test and effort expended on it, the relationship might be revealed in a greater tendency to use the available answers by the Valid-Test Ss.

Thus two variables were manipulated in this experiment: the purported validity of the test and the availability of the correct answers while taking it. A 2×2 factorial design was used, with 10 Ss in each treatment combination.

After the administration of the test, in which the time limit was reduced from the usual 15 minutes to nine, Ss were asked to rate the test "on the basis of the information you have about the test and the things you noticed while taking it." Four choices were available: a good valid test, a fairly good test, not too useful a test, and a worthless test. No further mention of the test was made after the materials were collected; then a week later the instructor announced that the tests would not be scored "because of the faulty nature of some of the test booklets used."

Results

Those 20 Ss who were led to believe that the test was valid were significantly higher in their ratings of the test's validity than were the 20 Ss who were led to believe the test was invalid. The difference, after dichotomizing the rating choices and using a chi-square test, is significant beyond the .01 level. Thus it appears that the answer-sheet information was effective in creating different perceptions of the validity of the test.

Mean scores for Ss in the four treatment combinations were as follows:

Valid Test--Answers Available	24.9
Valid Test--Answers Unavailable	20.6
Invalid Test--Answers Available	23.6
Invalid Test--Answers Unavailable	21.2

A 2×2 factorial analysis of variance was computed, with the results presented in Table I. An F-value of 6.74 was found for the availability-of-answers treatment. With 1 and 36 degrees of freedom, this is statistically significant at the .05 level. A t-test revealed that those Ss who had the correct answers available received significantly higher scores on the average (24.25) than did those Ss without answers available (20.9). This difference of 3.35 points was significant beyond the .05 level. The F-values for the test-validity treatment and for the interaction between the two variables did not approach statistical significance. The mean scores for the valid-test and invalid-test conditions were 22.75 and 22.4, respectively. Thus the purported validity of the test did not influence the average score, nor did Ss under the different validity conditions differentially utilize the available answers.

Discussion

The finding that students with correct answers available do better on an intelligence test than do those without is not particularly unexpected. Of more interest is the finding that there was apparently about an equal use of the answers regardless of whether the test was considered valid or not, and any differences in motivation caused by the validity of the test were not revealed in level of test performance. Apparently the answers were utilized by some Ss even in a situation where the only purpose of taking the test (in addition to its requirement for the course) was to give the student information about himself. When students are in conflict as to whether to obtain an accurate self-evaluation or the best possible score--even using information that they would not ordinarily have--some are unable to avoid seeking an unrealistically favorable evaluation of themselves.

That the purported validity of an intelligence test does not affect the average score on it is in conflict with the findinga of the achievement-motivation researchers (7). However, these previous researches used different kinds of tasks (less complex), different kinds of instructions (usually the importance of the S's performance was varied), and sometimes different kinds of Ss (both college students and Air Force recruits were used). For example, French (4) manipulated the importance of coding tasks by telling the recruits that the tasks were new ones, being tried out, or that they were important in determining one's future career. Despite their accurate understanding of the validity of the test, the Ss in the present study may not have felt the intelligence test important enough to exert a great deal of effort. Thus a tendency toward apathy in college Ss in such situations might have caused the lack of differences. On the other hand, the general "test-wiseness" of college students may have developed in them a general drive to do well on tests, regardless of the test's nature or validity. It is apparent that these explanations may not be verified until further research is done.

The work of French (4) indicates that a person's "typical" level of achievement motivation (i. e., not experimentally aroused) may influence his responsiveness to the experimental manipulation of the importance of a task. As no measures of achievement motivation are available for these Ss, it is impossible to know whether the effects of such differences have been glossed over in this experiment.

It does seem, on the basis of this study, that any generalization that the experimental manipulation of testing instructions and test characteristics will necessarily result in differences in performance is unwarranted. Likewise, perhaps some of the concern for the importance of face validity, at least for college students, is unwarranted.

REFERENCES

1. Clark, R. A., and McClelland, D. C. "The Effects of Sexual Motivation on Phantasy." In D. C. McClelland (Ed.), Studies in Motivation, New York: Appleton-Century-Crofts, 1955.
2. Cronbach, L. J. Educational Psychology, New York: Harcourt Brace, 1954.
3. Downie, N. M. Fundamentals of Measurement, New York: Oxford University Press, 1958.
4. French, Elizabeth G. "Some Characteristics of Achievement Motivation," In J. W. Atkinson (Ed.), Motives in Fantasy, Action and Society, Princeton, N. J.: Van Nostrand, 1958.
5. Garrett, H. E. Testing for Teachers. New York: American Book Company, 1959.
6. Lowell, E. L. "The Effect of Need for Achievement on Learning and Speed of Performance," Journal of Psychology, 1952, XXXIII, 31-40.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. The Achievement Motive, New York: Appleton-Century-Crofts, 1953.
8. Noll, V. H. Introduction to Educational Measurement, Boston: Houghton-Mifflin, 1957.
9. Pressey, S. L., Robinson, F. P., and Horrocks, J. E. Psychology in Education, New York: Harper, 1959.
10. Tiffin, J., and Lawshe, C. H. "Examiner Manual for the Adaptability Test," Chicago: Science Research Associates, 1954.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 4, December 1960)

INTERCORRELATIONS OF THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION WITH OTHER TYPES OF TESTS

J. C. GOWAN
San Fernando Valley State College
Northridge, California

AN IMPORTANT question for anyone who wishes to be an intelligent consumer of tests is the extent to which they correlate with standard instruments for measuring intelligence. It is important to know what variance the intelligence component is contributing to a given personality measure so that its influence may be properly assessed. The American Council Psychological Examination, college form, has had wide use as a group criterion of intelligence in young college adults. To minimize the relationship between intelligence and other variables, a relatively homogeneous population consisting of upper division college teaching candidates in a large state university was selected. Information on this population is elsewhere available (3),and indicates that it is about two-thirds women. Intercorrelations were secured by subjecting IBM sortings to Flanagan's tails method (1), which, experience elsewhere reported (2) indicates, results in standard errors about twice the size of those obtained by the Pearson product-moment method.

Table I presents the means, standard deviations, and standard errors of the means for various tests. It also gives the correlations between the Q and L sub-scores and the total score of the American Council Psychological Examination (hereafter called the ACE) with itself, the Cooperative English Test, the Stanford Arithmetic Test, the Allport-Vernon Study of Values and the Minnesota Multiphasic Personality Inventory. There is correlation between L and ACE total practically to the level of reliability of the test itself. The Cooperative English and the Stanford Achievement show high correlations with the ACE total, and differential ones with the Q and L subscores as might be expected. Other correlations with the criterion at the 1% level of significance are the Allport theoretical and aesthetic scales (positively), the economic and the MMPI lie scale (negatively).

Table II presents the same statistics for various other tests. Of the special aptitude tests only the Cardall, the MacQuarrie and the Morris Trait Index L (leadership) correlate significantly. There are no significant correlations with the Seashore, and only with the scientific interest in the case of the Kuder. Highly significant correlations, however, are found between the ACE and the Cooperative General Culture Test in total score and all subscores except Fine Arts. The same is true with respect to the Iowa Reading Test. Among the personality tests the ACE correlates positively with the T score of the Minnesota T-S-E and negatively with the E scale. This would indicate that thinking introversion and emotional introversion are differentially associated with intelligence. There are no significant correlations with the Bernreuter, but strong significant negative correlations in the case of the Bell total and the Emotional subscale.

In general it may be noted that basic skills (reading and arithmetic) correlate highest with the ACE. Next come achievement tests in content fields. Lower still but significant in some cases are personality measures of a desirable nature. Lowest and barely significant in but a few cases come special aptitudes and interests, such as musical, mechanical, artistic and the like. Implications for the counselor may be stated as follows: 1) expect moderate to severe contamination from intelligence in tests of basic skills and achievement, 2) recognize the significant relationships between intelligence and certain personality variables. In part this means that intelligence can be described in part as an activity associated with thoughtful, restrained, well-adjusted, theoretical, flexible, thinking-introverted, cultivated affect and the habits and behaviors which go with it. 3) recognize the essential independence of special aptitudes and interests from verbal intelligence.

---

1. Flanagan, J. C. "The Effectiveness of Short Methods of Calculating Correlation Coefficients," Psychological Bulletin, 49:4:342-348, (July, 1952).
2. Gowan, J. C. "Intercorrelations and Factor Analysis of Tests Given Teaching Candidates," Journal of Experimental Education, 27:1; 1-22, (September, 1958).
3. MacLean, M. S. et. al. "A Teacher Selection and Counseling Service," Journal of Educational Research. 48:9:669-678, (May, 1955).

TABLE I

BASIC STATISTICS AND CORRELATIONS WITH AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION OF COOPERATIVE ENGLISH, STANFORD ARITHMETIC, COOPERATIVE ENGLISH, ALLPORT-VERNON, AND MMPI

Test and Scale	mean	sigma	S.E.m	Correlation with ACE		
				r(Q)	r(L)	r(T)
A.C.E.						
Q(quantitative)	46.5	10.5	.25	---	.48**	.78**
L(linguistic)	78.8	16.0	.38	.48**	---	.92**
T(total)	125.4	22.9	.55	.78**	.92**	---
Coop. English						
Vocabulary	66.3	9.2	.22	.19**	.78**	.65**
Speed	63.3	10.7	.26	.48**	.78**	.77**
Level	63.7	9.2	.22	.34**	.65**	.60**
Mechanics	58.8	9.6	.23	.36**	.56**	.56**
Effectiveness	60.7	8.8	.24	.39**	.61**	.61**
Stanford Arith.	81.0	10.9	.26	.68**	.44**	.61**
Allport Values						
Theoretical	40.1	8.0	.19	.09	.23**	.20**
Economic	46.9	8.0	.19	.03	-.24**	-.16**
Aesthetic	41.7	9.2	.22	-.04	.22**	.16**
Social	40.8	7.0	.17	-.02	.00	.01
Political	39.6	6.8	.16	.01	-.06	-.04
Religious	39.4	10.3	.25	.04	-.15*	-.14
MMPI						
L	4.1	2.2	.05	-.11*	-.20**	-.20**
F	3.0	2.1	.05	-.07	-.06	-.02
K	18.7	4.4	.11	.04	.02	.01
Hs	12.7	3.0	.07	-.04	-.10*	-.10*
D	18.5	3.9	.09	-.14*	-.02	-.08
Hy	21.4	4.1	.10	-.07	-.08	-.08
Pd	21.5	4.0	.10	-.08	-.01	-.05
Pa	9.1	2.3	.06	-.08	-.06	-.09
Pt	25.9	4.4	.11	-.10*	-.04	-.08
Sc	25.7	4.7	.11	-.05	.03	-.01
Ma	19.1	3.8	.09	-.01	-.04	-.01
Si	20.6	7.6	.19	-.05	.00	.00

* Significant at the 5% level
**Significant at the 1% level

N varies from 1700 to 1732; 62% were women. Maximum S.E. of r when product-moment correlations are used would be .024. These correlations computed by Flanagan's tails methods. Experience indicates conservative estimate for maximum S.E.r under these conditions would be .05, which is figure used.

TABLE II

BASIC STATISTICS ON FURTHER TESTS AND CORRELATIONS WITH INTELLIGENCE

					Correlation with ACE (T)	
Test and Scale	N	mean	sigma	S. E. m	r	S. E. r
Cardall Judgment	100	202.6	24.4	2.44	.34**	.09
Minn. P. Fm. Bd.	100	46.0	8.40	.84	.15	.10
MacQuarrie	100	73.9	11.76	1.17	.27**	.09
Meier Art Judgment	100	100.7	7.96	.80	.19	.10
Seashore Music(tot)	100	220.7	13.15	1.32	.00	.10
1. Pitch	100	46.1	4.68	.47	.12	.10
2. Loudness	100	44.8	4.18	.42	.12	.10
3. Rhythm	100	27.6	2.46	.25	-.01	.10
4. Time	100	40.0	4.10	.41	.00	.10
5. Timbre	100	39.5	5.02	.50	-.11	.10
6. Tonal Memory	100	27.4	3.37	.34	-.05	.10
Kuder Preference R.						
1. Mechanical	100	51.7	16.4	1.64	.01	.10
2. Computational	100	26.1	13.7	1.37	.11	.10
3. Scientific	100	50.5	14.6	1.46	.21*	.10
4. Persuasive	100	69.1	15.7	1.57	-.12	.10
5. Artistic	100	57.1	16.9	1.69	-.08	.10
6. Literary	100	53.4	14.6	1.46	.09	.10
7. Musical	100	24.7	9.4	.94	.00	.10
8. Social Service	100	90.7	18.4	1.84	-.05	.10
9. Clerical	100	47.5	15.2	1.52	-.10	.10
Morris Trait Index	131	52.6	6.06	.61	.41**	.07
Coop. Gen'l. Cult.	100	180.9	44.5	4.45	.40**	.08
Current Events	100	42.9	9.5	.95	.36**	.09
History	100	33.5	12.8	1.28	.41**	.08
Literature	100	32.5	13.5	1.35	.41**	.08
Science	100	22.7	10.0	1.00	.20*	.10
Fine Arts	100	37.5	14.7	1.47	.19	.10
Mathematics	100	16.5	10.2	1.02	.29**	.09
Iowa Reading(med.)	94	187.6	13.9	1.43	.42**	.08
Minn. T-S-E						
T (Thinking)	118	113.3	22.2	2.04	.30**	.08
S (Social)	111	134.6	20.6	1.94	.00	.10
E (Emotional)	114	58.3	11.15	1.04	-.18*	.09
Bernreuter						
FIC (Self-conf.)	97	-67.4	83.2	8.40	-.12	.10
F2S (Sociability)	88	-35.4	46.4	4.92	.16	.11
Bell Adj. (total)	125	26.5	15.8	1.42	-.23**	.08
Home	125	5.62	5.46	.48	-.13	.09
Health	125	5.17	3.60	.32	-.16	.09
Social	125	7.90	6.40	.57	-.12	.09
Emotional	250	6.95	5.68	.36	-.32**	.06

Pearson product-moment correlation coefficients used

* significant at 5% level

**significant at 1% level

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 JANUARY, 1961 NUMBER 5

CONTENTS

Development of the Ideal Self in Western Australian Youth—D. K. Wheeler 163

Actuarial Prediction in Guidance—Henry D. Rinsland 168

Factor Analyses of High School Science Achievement Measures—Leroy Wolins, A. C. MacKinney and Paul Stephans 173

Concepts of Meaning and the Reading and Spelling Difficulty of Words—Richard H. Bloomer 178

Values and Achievement in a Freshman Psychology Course—William D. Winter 183

A Re-evaluation of the Heath Railwalking Test—Cornelius P. Goetzinger 187

Student Mortality and Related Factors—Norman Gekoski and Solomon Schwartz 192

A Comparison of the WISC Patterns of Male Retarded and Non-Retarded Readers—Donald Neville 195

A Study of Academic Achievement and Performance on a Test of Library Understandings—William D. Joyce 198

⁂

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

DEVELOPMENT OF THE IDEAL SELF IN WESTERN AUSTRALIAN YOUTH

D. K. WHEELER
University of Western Australia
Nedlands, Western Australia

HAVIGHURST, Robinson and Dorr[1]* have suggested a developmental trend in the growth of the ego-ideal or ideal self. From childhood identification with a parental figure it progresses, in preadolescence and early adolescence, through a stage of "romanticism and glamour", culminating in later adolescence in a stage which may be symbolized by an attractive, known, young adult, or seen as a composite or imaginary figure. Havighurst and Macdonald[2] have repeated the original American work with a larger and more representative group of New Zealand children and have found the same developmental trend. This paper presents the results for a sample of Western Australian adolescents still at school.

Havighurst suggests that when boys and girls are asked to write a brief essay on the topic "The Person I Would Like to Be Like," the persons written about can be placed in the following categories:

P Parents.

S Parent surrogates: grandparents, teachers.

G Glamorous adults with romantic or ephemeral fame due to the more superficial qualities of behavior and appearance: e.g., athletes, military figures, stage and screen stars.

H Heroes.

St. Saints or religious figures.

A Attractive adults known to the child: usually young neighbors, siblings or relations, youth leaders.

C Composite or imaginary characters. These are abstractions of a number of people, either wholly imaginary or a coalescence of qualities seen in two or three real persons.

M Age mates.

O Occupation only is stated, and so responses are not classifiable in the preceding categories.

NC Responses not classifiable elsewhere.

Procedure

As part of a long, group-administered questionnaire used to investigate the attitudes and opinions of adolescents still at school, students were asked to write a brief essay about "The Person I Would Like to Be Like." The last item on this questionnaire read:

"161 Read these instructions first
"The Person I Would Like to Be Like"
Describe briefly the person or sort of person you would most like to be like when you grow up.
This may be a real person or an imaginary person. He or she may be a combination of several people. Tell something about this person's age, character, occupation, and so on. If he or she is a real person, say so, but you needn't give the real name if you do not want to.
Go straight ahead."

These instructions are almost the same as in the original American study, though the wording is slightly different from that used in New Zealand.

With the exception of children in schools in the more remote parts of Western Australia, the questionnaire which concluded with this item was presented to children by the one person[3] who assured tham that their replies would be treated as confidential and would not be seen by their teachers or parents. Teachers were never present in the room while children were filling out the items. No names were required on the questionnaire, though students could sign them if they so desired.

Data were secured from children in all types of schools in Western Australia. The population from which the sample was selected included all boys and girls undergoing full-time education at the first, third, and fifth years of secondary schooling. The sample was stratified on the basis of a) sex composition of the population at school, b) proportion of urban to rural schools, and c) school type—government schools, private colleges, and Roman Catholic schools. The third year sample also included girls attending business colleges full time, as they were working toward the Junior Public Examination taken by most fifteen-year-olds in other schools. The sample drawn consisted of 300 first-year, 294 third-year, and 190 fifth-year boys and girls in school, comprising 4%, 10% and 24% of the school population in these years. Owing to the fact that this was the last item of a very long questionnaire, some students did not answer it. The number of replies analyzed and usable totalled 276 first year, (mean age 13 years and 5 months), 257 third year (mean age 15 years 4 months) and 166 fifth year (mean age 17 years 2 months).

---

* Footnotes will be found at the end of the article.

These groups are representative samples of the school populations at the three levels, but it should be noted that they contain no adolescents who have left school.

Results

The Western Australian data are summarized in Table I, using the same categories as in the U. S. and New Zealand investigations. The same general features are to be found.

a) With increasing age, there is a decrease in the number of glamorous persons described.

b) With increasing age, there is an increase in the number of composite or imaginary figures described.

The main comparisons which can be made are within the Western Australian sample, and between the New Zealand and Western Australian students at ages thirteen and fifteen. The Western Australian data can also be compared with some aspects of the American results.

Within the Western Australian sample there are no differences between boys and girls at ages 13 and 15. In the 17-year-old group, however, there is a sex difference significant at the one percent level. More girls than boys chose parental figures and attractive adults, and more boys than girls wrote about occupation only.

The comparisons between the Western Australian and New Zealand[4] samples are shown in Table II. The boys in the two countries do not differ significantly, but both the 13- and 15-year-old girls show differences significant at the 1% level. Western Australian girls chose more parental figures and fewer attractive adults.

From the United States figures given in Havighurst and Macdonald's Table IV, it is possible to make comparisons between United States and Western Australian youth. The 13- and 15-year-old groups for Western Australia are compared with three American subgroups in combination. These comparisons are shown in Table III, but it should be remembered that the groups are only roughly equivalent and the results should be treated with caution. There are no significant differences among the girls, but Western Australian boys choose significantly more composite and imaginary figures than United States boys. They also choose fewer parents and glamorous figures, though the differences are not significant.

All the data from the three countries show a tendency with increasing age to choose composite or imaginary rather than specific persons. To investigate this further, all choices of specific persons (P, S, G, H, St, A and M) were combined and contrasted with composite or imaginary figures. Among Western Australian boys there is a significant diminution of the number of specific persons and an increase in composites.· The difference between 13- and 15-year-olds is significant at the five percent level and that between 15- and 17-year-olds at the one percent level. Although the trend for girls is the same, the differences are not so marked. There are no significant differences between the 13- and 15-year-old groups, or between the 15- and 17-year-old groups, but that between the 13-year-olds and the 17-year-olds is significant at the one percent level.

When this same comparison is made between Western Australian and New Zealand 13- and 15-year-olds, there is no significant difference between the two samples, for either sex.

If, on this same basis, Western Australian 13-15-year-old boys are compared with United States 12-16-year-olds, there is a difference significant at the two percent level. As was seen before, the Western Australian boys choose more composites and fewer specific persons. The same tendency is present in the girls, but the differences are not significant.

Discussion of Results

Age Progression. The Western Australian data agree in many respects with those from the other two countries. There is, with age, a progression toward choosing a composite or imaginary character rather than a specific person. Boys and girls are alike in this respect at the lower age levels, but there are sex differences among the 17-year-olds, though the same trend is maintained. The United States and the New Zealand data are not concerned with 17-year-olds, so that there is no means of finding out whether the sex differences are peculiar to this country or not.

It will be seen that at 17, more girls chose parental figures and more boys wrote about occupations, so that their answers were not classifiable in the other categories. It is possible to suggest that the increase in the number of boys writing about occupations is due to the fact that they are in their last year at school and that their future occupation is looming large in their thoughts at this time. Table I indicates that there is an increase in the Occupation category from 13 through 17. The youngest group have at least another year at school, many of the 15-year-olds will leave at the end of the year, and all the 17-year-olds will do so. Or the explanation may be simpler still. This was the last question on a very long questionnaire, and those who say, "I would rather be a motor mechanic," or who enlarge on occupational characteristics, may simply be taking the shortest way of answering the question.

The increase in girls' choice of parental figures with increasing age is of interest. At 17 years, al-

TABLE I

CLASSIFICATION OF PERSONS DESCRIBED AS THE IDEAL SELF BY WESTERN AUSTRALIAN SCHOOL ADOLESCENTS

	Boys			Girls		
Age Group	13	15	17	13	15	17
Number of Papers	137	117	93	139	140	73
Category	Percentage of Responses					
P	4	8	3	7	9	11
S	--	2	2	2	5	1
G	21	8	4	16	10	4
H	1	--	--	3	1	--
St.	--	--	--	1	--	1
A	20	20	4	21	18	14
C	36	50	62	38	45	58
M	5	2	2	3	1	1
O	4	7	12	3	5	1
NC	8	3	9	6	5	8

TABLE II

COMPARISON OF WESTERN AUSTRALIAN AND NEW SEALAND YOUTH

		Percentage Frequencies			
		Age 13		Age 15	
		W. A.	N. Z.	W. A.	N. Z.
	Category				
Boys	P	4	9	10	5
	G	21	22	8	14
	A	20	25	20	24
	C	36	35	50	48
	N. C.	8	3	3	4
	Number of Subjects	137	179	117	111
Girls	P	9	1	14	2
	G	16	17	10	9
	A	21	38	18	26
	C	38	39	45	52
	N. C.	6	---	5	4
	Number of Subjects	139	177	140	142

TABLE III

COMPARISON OF WESTERN AUSTRALIAN AND AMERICAN YOUTH

Category	Percentage Frequencies			
	Boys		Girls	
	U. S., Age 13-16	W. A., Age 13-15	U. S., Age 13-16	W. A., Age 13-15
P	13	7	11	11
G	23	15	14	13
A	23	20	25	20
C	30	42	37	42
Number of Subjects	248	254	280	279

most all of the girls who choose parental figures are away from home, either at boarding school or living in school hostels, in lodgings or with friends or relations. This may be pure coincidence, or it may be that girls away from home tend to keep the parents as an ego-ideal for a longer period. Certainly in boarding schools and school hostels away from the family home, there is less chance of a girl coming into contact with attractive young adults.

The Western Australian sample throughout includes a higher proportion of non-classifiable answers than the New Zealand replies. Almost all of these say, in effect, "I don't want to be like anybody else. I just want to be as I am." Havighurst et al. (1, p. 257) have suggested that this may arise from unconscious resistance to recognition of the nature of the ideal self. This may not be so, or these respondents, having already answered so many questions, may take the shortest and easiest way of satisfying the investigator without going to the length of refusing to answer.

Comparisons Between Countries. American boys differ from both Western Australian and New Zealand boys, who seem much alike. Western Australian girls' groups seem more similar to American groups, though it has already been suggested that this comparison is a very tentative one in view of the fact that the American data may not be representative of American girls and was collected a decade prior to the Western Australian investigation.

Relative Maturity. If choice of a composite or imaginary character be taken as a sign of greater maturity, it can be concluded that Western Australian and New Zealand youth (to age 15) progress at about the same rate. This representative sample of Western Australian youth would seem to display somewhat greater maturity than the earlier, not necessarily representative, sample of American youth. This conclusion is advanced tentatively, subject to further American work.

Qualities of the Ideal Self. Analysis of the replies categorized as Composite or Imaginary indicate that many of the characteristics of the persons described are ones which have already been mentioned in other questions (in the total survey). The 13- and 15-year-old groups make a good deal of mention of the body image, with details about the sort of build, face, figure and features that these boys and girls admire. The oldest seem more accepting of their physiques, though the girls tend to stress makeup, dress, and hair styles. Most of the composite or imaginary figures are good at sport, popular, able to mix well with other people, and make friends. Many of the boys and girls mention that they want to have a wife or husband, be happily married, have a specified number of children, a comfortable home and a good job. These common characteristics of the ego-ideal are very similar to the information obtained from other parts of the questionnaire[5].

Conclusions

This investigation shows that there is present in Western Australian youth, as in New Zealand and United States youth, the same developmental trend in the ideal self. From early to late adolescence, there is a diminution in the number of parental figures chosen and an increase in the number of characters who are either imaginary composites of desirable qualities or blends of admired traits abstracted from more than one real person. The evidence is insufficient to support Havighurst's thesis[6] completely, but it does show clearly the change with age from the parental image to the composite figure.

FOOTNOTES

1. Havighurst, R. J.; Robinson, Myra Z., and Dorr, Mildred, "The Development of the Ideal Self in Childhood and Adolescence," Journal of Educational Research, XL (December, 1946), 241-257.
2. Havighurst, R. J. and Macdonald, D. V., "Development of the Ideal Self in New Zealand and American Children," Journal of Educational Research, XLIX (December, 1955), 263-273.
3. Julie M. Wheeler, Senior Research Fellow, Faculty of Education, University of Western Australia.
4. The New Zealand figures are taken from Table III, Havighurst and Macdonald (ibid. p. 270).
5. Wheeler, Julie M., "Popularity and Status in the Adolescent Peer Group," Educand 1, 4, 1953, pp. 80-91.
6. Havighurst and Macdonald (ibid.) p. 273.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

ACTUARIAL PREDICTIONS IN GUIDANCE[*]

HENRY D. RINSLAND
University of Oklahoma
Norman, Oklahoma

THE SETTING of this paper is in harmony with the major theme of this program of the AAAS--the emergency in the training and procurement of scientists. Actuarial predictions of potential scientists must take their setting in a full program of guidance with the following classifications:

1. A knowledge of vocations, their supply and demand, and their necessary educational preparations are called educational-vocational guidance and which is a continuum.
2. Detailed knowledge of the individual, his capacities and the success of his efforts in learnings, are called clinical guidance.
3. The development of personality, culture, and citizenship are called cultural guidance.
4. A prediction between the individual's abilities and his achievements both in college and vocation, I shall call actuarial, statistical, or predictive guidance.

That the knowledge of the supply and demand of the vocational world is terrifically important today is a universal cry all over America. We are not now producing the scientists, mathematicians, nor the school teachers needed in this age of ever demanding educational qualifications. But the training of competent workers in any field first demands identification before the training begins. But more valid and accurate measurements must be used for the best prediction we make of their success in advanced training. The education of future workers in respective fields is not simple and every appraising instrument used should possess the highest validity and reliability obtainable. Standardized tests of general ability and specific aptitudes have been produced which are many times more reliable than high school grades but these are the ultimate criteria of successful training in any profession.

And yet with this more definite test-knowledge of the individuals, the colleges know little about the predictions from tests or school grades to college grades as shown by the American Council on Education Committee report.[1**] Here, as in all other studies, the real tragedy is the low reliability of the criterion which are college grades. So little is known about this important criterion that the Stuit committee estimated the reliability of college grades, seldom as high as .55 and, perhaps, average around .40 or .50. Literally thousands of correlations are published showing relationship of high school measurements and grades with college achievement but no where are these correlations interpreted in term of chances of success; yet this type of prediction is exactly what the counselor and the students need. It is the only kind of prediction that really makes sense to all concerned.

But high school and college grades which are important criteria in prediction can be made more reliable. That those techniques are workable on a large scale is shown by the Air Command.[2] The simple device to secure reliable grades in any school is to first build reliable and valid objective and subjective subject-matter tests, convert their cumulative scores into a mathematically defined and calculable letter grade by what is sometimes called the Sigma Plan where each letter grade is defined as one standard deviation wide.

With increased reliabilities and validities, all subject-matter grades used as criteria of all kinds of predictions (basic subject matter, aptitude and general intelligence) can be used for two things: first, increase the correlation between tests and grades, and second, increase the validity of tests themselves where grades are used as criteria.

In order to show the value of these predictions, and to actually use statistical predictions from high school tests or grades to college tests or grades and/or from college grades to vocational success, these are presented from simple regression equations based upon normal probabilities, twelve tables of predictions from a correlation (r) of .40 to .95 by steps of .05. These are also applicable to multiple R's (aside from a few qualifications) where many variables are combined and correlated with a criterion.[3] The regression equation is based upon the correlation between two variables and the higher the correlation the more accurate the prediction from

---

[*] A condensation of a speech delivered at Section Q, Education, A.A.A.S., Atlanta, Georgia, 1955.

[**] All footnotes will be found at end of article.

TABLE I

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .40

X Y	A	B	C	D	E	Total
A	15	27	20	5	0	67
B	27	82	93	35	5	242
C	20	93	157	93	20	383
D	5	35	93	82	27	242
E	0	5	20	27	15	67
Total	67	242	383	242	67	1001

TABLE II

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .45

X Y	A	B	C	D	E	Total
A	17	28	18	4	0	67
B	28	87	92	31	4	242
C	18	92	163	92	18	383
D	4	31	92	87	28	242
E	0	4	18	28	17	67
Total	67	242	383	242	67	1001

TABLE III

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .50

X Y	A	B	C	D	E	Total
A	19	29	16	3	0	67
B	29	91	92	27	3	242
C	16	92	167	92	16	383
D	3	27	92	91	29	242
E	0	3	16	29	19	67
Total	67	242	383	242	67	1001

TABLE IV

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .55

X Y	A	B	C	D	E	Total
A	21	30	14	2	0	67
B	30	96	91	23	2	242
C	14	91	173	91	14	383
D	2	23	91	96	30	242
E	0	2	14	30	21	67
Total	67	242	383	242	67	1001

TABLE V

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .60

X Y	A	B	C	D	E	Total
A	23	30	13	1	0	67
B	30	97	91	23	1	242
C	13	91	175	91	13	383
D	1	23	91	97	30	242
E	0	1	13	30	23	67
Total	67	242	383	242	67	1001

TABLE VI

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .65

X Y	A	B	C	D	E	Total
A	26	31	9	1	0	67
B	31	108	87	15	1	242
C	9	87	191	87	9	383
D	1	15	87	108	31	242
E	0	1	9	31	26	67
Total	67	242	383	242	67	1001

TABLE VII

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .70

X Y	A	B	C	D	E	Total
A	30	30	7	0	0	67
B	30	116	85	11	0	242
C	7	85	199	85	7	383
D	0	11	85	116	30	242
E	0	0	7	30	30	67
Total	67	242	383	242	67	1001

TABLE VIII

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .75

X Y	A	B	C	D	E	Total
A	34	29	4	0	0	67
B	29	126	80	7	0	242
C	4	80	215	80	4	383
D	0	7	80	126	29	242
E	0	0	4	29	34	67
Total	67	242	383	242	67	1001

TABLE IX

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .80

Y \ X	A	B	C	D	E	Total
A	38	27	2	0	0	67
B	27	137	74	4	0	242
C	2	74	231	74	2	383
D	0	4	74	137	27	242
E	0	0	2	27	38	67
Total	67	242	383	242	67	1001

TABLE X

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .85

Y \ X	A	B	C	D	E	Total
A	43	23	1	0	0	67
B	23	154	63	2	0	242
C	1	63	255	63	1	383
D	0	2	63	154	23	242
E	0	0	1	23	43	67
Total	67	242	383	242	67	1001

TABLE XI

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .90

Y \ X	A	B	C	D	E	Total
A	50	17	0	0	0	67
B	17	178	47	0	0	242
C	0	47	289	47	0	383
D	0	0	47	178	17	242
E	0	0	0	17	50	67
Total	67	242	383	242	67	1001

TABLE XII

NUMBER OF CASES OUT OF ONE THOUSAND FOR EACH PENTAD OF GRADES (X) PREDICTED FROM EACH NORMALIZED PENTAD OF A TEST (Y) FOR A CORRELATION OF .95

Y \ X	A	B	C	D	E	Total
A	60	7	0	0	0	67
B	7	215	20	0	0	242
C	0	20	343	20	0	383
D	0	0	20	215	7	242
E	0	0	0	7	60	67
Total	67	242	383	242	67	1001

one variable to another.

This element of prediction in guidance is essential in both guidance to college and to vocations. Without such real predictions guidance is about at the level of personal judgment and advice, or with expert psychologists, somewhat better; just how good has not been determined from experimental or statistical evidence.

Meehl[4] has presented arguments of clinical and statistical preductions, but statistical predictions vary with correlations. The ultimate criterion of all predictions is, of course, a percentage of cases which actually achieve the predicted achievement and this is a statistical evaluation.

The most widely used prediction is probably between test scores and grades. For grades there are usually five steps or pentads[5] which should be equal in linear units (like inches) and for all practical purposes each 1 standard deviation wide (from -2.5σ to $+2.5\sigma$). Similar divisions could be made from test scores and, if not given in the test manual, as such, can be calculated with little trouble as the means and standard deviations are always published. These divisions then permit direct comparison of teachers' marks.

Jackson[6] has suggested that by using standard scores predictive percentages may be read directly from a table of normal areas for any given correlation. A comment or two is in order to make clear the construction of these prediction tables since all data are based upon the assumption of theoretical frequencies of a normal bivariate surface.

It is necessary first to reduce both the regression equation and the standard error of estimate to a standard score form where $Y=rX$ and $S.E._{(est)}$ is equal to $S.D.\ \sqrt{1-r^2}$. The second point of observation is to read from a table of areas of the normal curve the percentage of cases that will fall within the following respectively named class intervals of 1 S.D. width: E, from -2.50σ to -1.49σ; D, from -1.50σ to $-.49\sigma$; C, from $-.50\sigma$ to $+.49\sigma$; B, from $+.50\sigma$ to $+1.49\sigma$; A, from $+1.50\sigma$ to $+2.50\sigma$; where we find 67, 242, 383, 242, and 67 cases respectively for 1,000 total--strictly speaking 1,001 cases.

From a given letter pentad grade or score of A, B, C, D, or E, one can predict the chances of making an A, B, C, D, or E in any future work for any respective correlation given. For example: reading from the Table I for a correlation of .40, the chances for an A student in high school marks or on an aptitude test score of making A in college is 15 out of 67 cases, of making a B is 27 out of 67, of making a C is 20 out of 67, of making a D is 5 out of 67, and making an E is 0 out of 67. Further, examining Table XII for correlation of .95, a student making an A in high school or on an aptitude has 60 chances of making an A out of 67, 7 chances of making a B out of 67, and 0 chances of making a C, D, or E.

As we increase the reliability of college grades the predictive power of aptitude tests can be raised; their reliability, usually in the .90's, already greatly exceeds the reliability of college grades. Aptitude tests have about reached the limit of both reliability and predictability. There is little room for improvement in any phase of educational guidance through statistical predictions except by raising the reliability of school marks. Horst[7] has presented formulas for differential prediction, but at present these are not within the statistical knowledge of most teachers. The tables presented here are both valuable and simple to use. These should aid in interpreting the thousands of useful correlations already published.

FOOTNOTES

1. Stuit, Dewey B., Predictions in Professional Schools, American Council on Education, 1949.
2. Air Training Command, Improvement of Grading Practices for Air Command Schools, Number 50-4, 1952.
3. McNemar, Quinn, Psychological Statistics, John Wiley and Sons, 1955, p. 176.
4. Meehl, Paul E., Clinical Versus Statistical Predictions, University of Minnesota Press, 1954.
5. For brevity in a title and convenience of a term that is not used in statistics, the term "pentad" from the Greek, meaning five equal units, is used. Pentad is an old chemical term meaning "with a valency of five" (Hackh's Chemical Dictionary, Third Revised Edition by Grant, 1944, Blakiston Company, Philadelphia, p. 623). Each unit, pentad, in the present usage, is set at one standard deviation wide, except for the two ends which will take care of all the extreme cases which occasionally occur.
6. Jackson, R.W.B., Ontario College of Education, letter, September 8, 1950.
7. Horst, Paul, "A Technique on the Development of a Differential Prediction Battery," Psychological Monograph, Vol. 68: 1954, American Psychological Association; and "The Prediction of Academic Achievement in College," College of Education Record, Washington University, Vol. 22, March, 1956.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

FACTOR ANALYSES OF HIGH SCHOOL SCIENCE ACHIEVEMENT MEASURES

LEROY WOLINS AND A. C. MACKINNEY
Iowa State University
Ames, Iowa

PAUL STEPHANS
Riverside-Brookfield High School
Riverside, Illinois

Introduction

SCIENCE EDUCATION in general and science education in secondary schools in particular is receiving a great deal of popular attention. This attention appears to be directed toward strengthening science course offerings in high schools. It would appear, therefore, additional information is needed to spell out further the nature and content of high school science achievement. This study is an attempt to specify the factor content of high school science achievement in one high school in suburban Chicago, Illinois.

The purpose of this study, then, is to explore the factor structure, using a factor analysis technique, of achievement in high school biology, chemistry, and physics together with other measures of aptitude, interest, and achievement that might aid in the isolation and identification of this factor structure.

Sampling

Three samples of students were used in the present study. All had completed at least one semester of biology, chemistry, or physics at the Riverside-Brookfield High School, Riverside, Illinois. The biology sample (N=119) had completed two semesters of a biology sequence. The chemistry sample (N=75) had completed two semesters of a chemistry sequence. The physics sample (N=87) had completed one semester of physics. These three samples are not independent. Thirty-seven students were common to all three samples, and still others were common to any two samples.

Design

Since every student did not take all five science courses or the respective standardized achievement tests for a given science, three analyses were made. These were for the biology, chemistry, and physics samples.

The factor analyses were carried out by means of Thurstone's group centroid technique. Erroneous communality estimates were adjusted by Wherry's method (3). The factors were rotated orthogonally for meaningfulness.

Variables Studied

A total of 16 variables were included in this study. However, all were not available for the three samples. These differences are indicated in the listing below. Raw scores were used throughout except for the course grades.

1. Cooperative General Achievement Tests, Revised Series, Test II. (Co)*. A general high school science achievement test administered during the seventh high school semester. (Available on the chemistry and physics samples only).
2. California Achievement Test, Advanced Battery. (CA). A measure of general achievement administered during the latter part of the eighth grade. (Available on all three samples.
3. California Short-form Test of Mental Maturity, Verbal. (V). A measure of verbal aptitude administered during the latter part of

---

* Letters in parentheses refer to the abbreviations for the variables used in Table I.

TABLE I

INTERCORRELATIONS, RESIDUALS, FACTOR LOADINGS, MEANS, AND STANDARD DEVIATIONS FOR THE VARIABLES INDICATED. FOR DESIGNATION OF THE VARIABLES, SEE TEXT

		1 Co	2 CA	3 V	4 NV	5 AV	6 E	7 H	8 N	9 A	10 K	11 B_1	12 B_2	13 C_1	14 C_2	15 P	16 S	I	II	III	h^2	
1.	B	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Co
	C	X	33	21	35	28	22	24	-	48	38	-	-	05	31	-	36	44	54	06	49	
	P	X	09	04	24	26	10	16	-	-	47	-	-	-	-	25	29	21	61	-	42	
2.	B	-	X	66	79	70	69	46	42	-	-08	54	51	-	-	-	-04	96	00	-15	94	CA
	C	02	X	55	69	64	64	24	-	47	-16	-	-	50	56	-	-09	80	-10	17	68	
	P	04	X	50	70	62	66	24	-	-	-20	-	-	-	-	49	-13	85	00	-	72	
3.	B	-	03	X	50	48	46	31	35	-	-09	43	44	-	-	-	03	64	00	03	41	V
	C	04	09	X	41	47	38	17	-	37	-19	-	-	34	38	-	-16	52	-13	17	32	
	P	-03	02	X	34	44	38	18	-	-	-20	-	-	-	-	31	-20	57	-09	-	33	
4.	B	-	04	-01	X	70	56	49	47	-	00	46	51	-	-	-	08	74	10	03	56	NV
	C	-02	03	-02	X	66	62	30	-	44	-04	-	-	37	48	-	-02	83	01	00	69	
	P	-03	04	-09	X	66	62	35	-	-	-07	-	-	-	-	42	-02	78	18	-	64	
5.	B	-	-09	-06	03	X	66	55	52	-	-02	54	58	-	-	-	01	84	04	09	72	AV
	C	-10	01	07	00	X	67	34	-	36	-14	-	-	40	48	-	-11	80	06	-02	64	
	P	04	-04	00	03	X	62	34	-	-	-14	-	-	-	-	46	-15	78	10	-	62	
6.	B	-	-03	-02	-02	02	X	33	30	-	-17	53	49	-	-	-	-12	75	-15	00	58	E
	C	-01	-03	-09	-02	07	X	27	-	35	-24	-	-	55	57	-	-20	77	-23	21	69	
	P	-03	-02	05	00	00	X	31	-	-	-25	-	-	-	-	50	-23	81	-07	-	66	
7.	B	-	-01	-04	04	07	-03	X	40	-	25	51	52	-	-	-	24	53	25	24	40	H
	C	02	-08	-05	03	08	-04	X	-	30	09	-	-	37	47	-	01	33	10	36	25	
	P	-02	-08	-03	01	02	00	X	-	-	08	-	-	-	-	35	00	39	17	-	18	
8.	B	-	-06	-01	00	03	-06	-03	X	-	24	53	57	-	-	-	35	54	31	28	47	N
	C	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	
	P	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	
9.	B	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	A
	C	05	03	08	01	06	-05	-05	-	X	21	-	-	39	60	-	20	51	35	42	56	
	P	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	
10.	B	-	01	-04	-02	00	01	06	00	-	X	-04	01	-	-	-	66	-08	81	11	67	K
	C	-01	00	-03	04	-11	02	04	-	-02	X	-	-	-28	-08	-	64	-10	81	00	67	
	P	05	01	01	-01	-02	01	04	-	-	X	-	-	-	-	01	60	-25	78	-	67	

TABLE I (Continued)

	1 Co	2 CA	3 V	4 NV	5 AV	6 E	7 H	8 N	9 A	10 K	11 B_1	12 B_2	13 C_1	14 C_2	15 P	16 S	I	II	III	h^2	
11. B	-	01	00	-06	-05	04	05	05	-	-01	X	83	-	-	-	-04	64	-05	55	71	
C	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	B_1
P	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	
12. B	-	-02	01	-02	-03	00	01	04	-	-01	05	X	-	-	-	-04	65	00	66	86	
C	-	-	-	-	-	-	-	-	-	-		X	-	-	-	-	-	-	-	-	B_2
P	-	-	-	-	-	-	-	-	-	-		X	-	-	-	-	-	-	-	-	
13. B	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	
C	-01	01	-04	02	09	00	02	-	00	02	-	-	X	74	-	-23	43	-32	68	75	C_1
P	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	
14. B	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	
C	-01	01	-02	05	07	00	03	-	03	-01	-	-	00	X	-	-15	53	-02	74	83	C_2
P	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	
15. B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	
C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	P
P	-02	01	01	-07	-01	06	09	-	-	03	-	-	-	-	X	08	57	24	-	38	
16. B	-	-08	00	04	-03	-03	02	08	-	00	-03	-07	-	-	-	X	04	81	00	66	
C	03	06	-01	04	-09	04	00	-	03	07	-	-	08	-03	-	X	-08	69	-08	49	S
P	-04	06	01	03	-05	00	-03	-	-	01	-	-	-	-	-	X	-22	69	-	52	
$\bar{X}$ B	-	301	39	51	50	53	56	123	-	42	2.4	2.4	-	-	-	.57					
C	63	309	41	52	53	55	60	-	115	45	-	-	2.8	2.5	-	.75					
P	63	308	41	52	52	54	59	-	-	45	-	-	-	-	2.4	.76					
S B	-	35	7.0	8.0	10	9.1	10	12	-	13	.94	.91	-	-	-	.49					
C	6.7	32	6.3	8.0	10	8.8	9.0	-	13	13	-	-	.99	1.1	-	.43					
P	6.4	30	7.0	7.8	9.5	9.1	9.0	-	-	13	-	-	-	-	1.0	.43					

B = 119
C = 75
P = 87

the eighth grade. (Available on all 3 samples).

4. California Short-form Test of Mental Maturity, Non-verbal. (NV). A measure of nonverbal reasoning aptitude administered during the latter part of the eighth grade. (Available on all three samples).
5. California Short-form Test of Mental Maturity, Advanced Verbal. (AV). A measure of verbal aptitude administered during the latter part of the fifth high school semester. (Available on all three samples).
6. Cooperative English Test; Lower and Higher Levels, Test A, Mechanics of Expression. (E). A measure of English achievement administered during the latter half of the sixth high school semester. (Available on all three samples).
7. Cooperative American History Test. (H). A measure of history achievement administered during the latter part of the sixth high school semester. (Available on all three samples).
8. Nelson Biology Test: Evaluation and Adjustment Series. (N). An objective measure of biology achievement administered after taking two semesters of high school biology, usually in the latter part of the fourth high school semester. (Available on the biology sample only).
9. Anderson Chemistry Test: Evaluation and Adjustment Series. (A). An objective measure of chemistry achievement administered after taking two semesters of high school chemistry, usually in the latter part of the sixth high school semester. (Available on the chemistry sample only).
10. Kuder Preference Record (Vocational), Scientific Scale. (K). A measure of scientific interest usually administered during the second high school semester. (Available on all three samples).
11. First Semester Biology Grade. (B_1). (Available on the biology sample only).
12. Second Semester Biology Grade. (B_2). (Available on the biology sample only.
13. First Semester Chemistry Grade. (C_1). (Available on the chemistry sample only).
14. Second Semester Chemistry Grade. (C_2). (Available on the chemistry sample only).
15. Physics Grade. (P). (Available on the physics sample only).
16. Sex. (S). Males coded one; females coded zero. (Available on all three samples).

Results and Discussion

Three factors were identified for two of the three samples, biology and chemistry, and, for the physics sample, two factors were identified. The intercorrelations (above the diagonal), residuals (below the diagonal), factor loadings, means, and standard deviations are all presented in Table I. Primary attention should be directed to the results of the factor analyses, the factor loadings and communalities, presented in the last four columns of the table. Since the residuals are all below .11 we can be reasonably certain that all major relationships are accounted for by the two or three factors.

The first factors (the plural is used since the same factor was identified in all three samples), the General Intelligence Factors, are identified by loadings from all the variables except Sex and Kuder Scientific. The highest loading on these factors is from the California Achievement Test. These factors are repeatedly found in studies dealing with pencil and paper tests and/or school grades where the factors are rotated orthogonally. Since these factors will be sufficiently familiar to the reader, they will not be discussed further.

The second factors (again, the plural is used for the same reason) are identified by loadings from all the science achievement tests and the physics grade, as well as very high loadings of Sex and the Kuder Scientific Scale. In addition, the loadings of the Cooperative American History Test and the Cooperative English Test (negative) should be noted. A consideration of these loadings lead the authors to believe this to be a Male Interest-Achievement Factor. This interpretation generally follows what is known of sex differences in school achievement and is based in part on the fact that girls typically do better in verbal subjects (1, pp 660-661). This accounts for the negative loadings of the Cooperative English on these factors. Also, boys do slightly better at history (1, pp 660-661) which is in harmony with the history loadings on these factors.

The high loadings of Sex and Kuder on factor two are interpreted as meaning that, independent of intelligence, males learn more about science than females. It would appear to the authors, that the Kuder Scientific Scale, at least for these students, is actually less a measure of scientific interest and more a conventional masculinity-femininity measure. The fact that the Kuder loads slightly higher on this factor than Sex itself may be attributed to the Sex variable being dichotomous and the Kuder variable being continuous.

The negative loading of first semester Chemistry grade on these second factors is interpreted to mean that the teacher of this course graded the more proficient males lower than the less proficient females. Results similar to these have been found by other researchers (2). We speculate that one of the few controls the high school teacher has over rambunctious students is grades. Thus, the typically rowdy male is graded lower whereas the more controlled female is graded higher. This interpretation is further supported by the lack of loadings of other science grades on this factor. It should be emphasized that here, as below, we are assuming that the standardized achievement tests are more valid measures of achievement than teacher grades.

The third factors were identified from the biology and chemistry samples only. No third factor could be present in the physics sample because only one physics course grade was available and there was no achievement test in physics.

As with factor two we find small loadings from the English and History tests. However, in this instance these loadings are not so easily explained since Sex does not load on factor three. As a result the authors have not been able to make good sense of this phenomenon. However, the higher loadings of the course grades and the achievement tests in the specific science areas lead the authors to believe these to be Specific Achievement Factors.

Conclusions

1. Three factors in high school science achievement were identified: General Intelligence, Male Interest-Achievement, and Specific Science Achievement.

2. The Kuder Scientific Scale probably should not be used alone for counseling students in curricular choice since it appears to be more a masculinity-femininity measure than a measure of scientific interest.

3. Science course grades appear to contain a sizeable component of invalid variance which is probably due to the use of grades by high school teachers as a means of keeping order.

4. Since high school science grades appear to contain this invalid variance, more emphasis should be placed on the use of standardized achievement tests, instead of grades, for criterion purposes.

Summary

Sixteen aptitude and science achievement variables from high school science students were intercorrelated and factor analyzed. Three factors were identified: General Intelligence, Male Interest-Achievement, and Specific Science Achievement. Various implications of the study, including inferences about the use of grades as criteria and the use of the Kuder Scientific Scale as a counseling tool, were discussed.

REFERENCES

1. Anastasi, A. and Foley, J. Differential Psychology. New York: Macmillan, 1949.
2. Lentz, T. F. Sex differences in school marks with achievement test scores constant. School and Society, 1929, 29, 65-68.
3. Wherry, R. J. A new iterative method of correcting erroneous communality estimates in factor analysis. Psychometrika, 1949, 14, 231-241.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

CONCEPTS OF MEANING AND THE READING AND SPELLING DIFFICULTY OF WORDS

RICHARD H. BLOOMER
State University Teachers College
Geneseo, New York

NUMBERS OF experiments have indicated learning of nonsense material to be less efficient than learning of meaningful material. Marks and Ollie (7) used second, third, fourth, and fifth order approximations of sentences as compared with regular textual material. They found that as material became more meaningful, the ability of college students to learn the material was markedly increased. Meaning, or at least degree of familiarity, seems then to be a variable in learning to repeat written material. Meaning as considered in terms of communication is not a simple variable.

There are basically three points of view on the relation between the referent or significate of a symbol, and difficulty in the communication process. The first of these positions, multi-ordinality, is concerned with the accuracy of meaning. In essence it indicates that a word does not communicate well if the symbol does not have a specific significate. Thus in so far as the symbol will represent a greater number of significates the probability of confusion, and hence error, in communication is increased. In essence this view espoused by Korzybski (4) and Johnson (3) says that each word is a homonym of itself and the greater the number of homonyms that any given sign represents, the more likely there is to be confusion concerning its meaning. Zipf's argument (14) that the number of meanings of a word increase as a function of the frequency of use of the word mitigates against a similar conclusion for learning to read or spell words. These functions are based upon repetitions, and as the number of individual repetitions approximates the number of repetitions in the population of words, the greater the number of meanings the easier the word should be to learn.

The second position, concreteness or its inverse abstraction, indicates that the difficulty in communication lies in the tangibility of the significate. Thus, the lowest level of difficulty or abstraction, is a concrete significate which can be contacted by the senses. Words which are contacted by a smaller number of senses, or are defined only in terms of other words, are less concrete, and hence represent greater communication difficulty.

Lorge (5) used a five point scale measuring the child's familiarity with the significate and found that this predicted the readability of a passage. This conclusion, drawn from the area of communication, might be applied to learning language materials where specific meanings are required for comprehension of the language.

The third position is that familiarity with the symbol implies familiarity with the significate. Thus, those words which occur most frequently in the language will tend, all things being equal, to communicate more specifically than those words which do not occur in the language.

There has been considerable work done with frequency of occurrence, particularly in relation to the organization of children's material. The most significant work was perhaps done by Thorndike (10) (11) and Lorge and Thorndike (6) in the Teachers Word Books. Thorndike is basically responsible for the notion that generalized familiarity makes for ease in learning. Whether this notion was originally true or not cannot be tested within public school populations, for these children have been exposed to words in relation to their frequency of occurrence, and hence, because they have had greater learning exposure, to more frequent words. The work of Noble (8) seems to answer this question for adults. His sample tended to know the meanings of words which were familiar to them when meaning was measured in word associations.

Procedure

The sample. The sample for the study represented words selected at random from The New Iowa Spelling Scale by Green. (2). Words which were not given a grade equivalent for reading in Gates (1) were discarded. The sample consisted of 149 words.

The Criteria

1. The grade level at which fifty percent of the children spelled the word correctly was derived from The New Iowa Spelling Scale (2) by linear interpolation.

2. Reading difficulty of the words was taken from Gates (1). Gates had children read a word and then select a synonym for it from more frequent words. This required considerable reading skill since both the word and the defining words had to be read correctly and connected to one another. The grade level at which fifty percent of the children read and chose correctly was the reading criterion for the study.

The Variables

1. Multi-ordinality represented the number of meanings of the word given in Webster's New World Dictionary (12). The following rules were established for determining number of meanings:

a. The number of dictionary meanings for all uses of the word were counted.

b. Meanings marked "archaic," "obsolete," or "rare" were not counted.

c. When a word occurred in an idiomatic expression, the idiomatic form and the number of extra meanings listed for it were counted.

d. Colloquial meanings were counted.

e. When one word was defined as part of speech of the second, applicable meanings of the second were counted.

f. Special meanings for other countries (e.g. Scotland or Britain) were not counted.

g. Sub-meanings under a single meaning were not included. For example "gang: a group of people associated in some way" was counted. The sub-meanings (a. a group of workers under one foreman, b. an organized group of criminals, c. a squad of convicts at work, d. a group of children banded together for social reasons) were considered part of the original meaning and were not counted. The counting of such sub-divisions or generalized meanings, could not be adequately done and would yield numbers that were far too large to have practical use. The number of specific meanings may be in some way related to the frequency of occurrence of the words.

2. Concreteness of the referent was measured by the modal response of five graduate student judges of the number of senses which are customarily applied by people to the first dictionary meaning of the word. The senses which were judged included the five natural senses and two which might be described as developed senses; that is to say, they are learned in relation to stimuli but are not necessarily a function of the stimulus *per se*. The senses are listed below.

a. Sight: do you usually see the referent?

b. Smell: do you usually smell the referent?

c. Taste: does your usual contact with the referent include tasting it?

d. Hearing: do you usually or normally hear the referent?

e. Tactile: is the referent customarily felt or touched?

f. Kinesthetic: does the contact with the referent usually involve some specific body position or change in position?

g. Time: is the referent stable for a period of time?

The score for any given word, then, represented the number of these senses which were judged by the students as customarily applied to the first dictionary meaning.

3. The frequency of occurrence and the $\log_2$ of the frequency of occurrence was determined from the total frequency reported in Rinsland's A Basic Vocabulary of Elementary School Children (9).

Results

Table I indicates that multi-ordinality related to spelling difficulty and to reading difficulty. The relationship of the number of meanings to spelling difficulty is larger than the relationship of the number of meanings to reading difficulty. Further, these relationships are not in the expected direction; that is to say, number of meanings does not appear to act as a confusion, but rather it is an aid in both spelling and reading difficulty. These relationships are mild, indicating the variables are not of great importance.

Table I indicates negative relationships between the concreteness of the significate and the grade level at which fifty percent of the children learned to spell and read the word. The relationship between concreteness and reading difficulty is stronger than the relationship between concreteness and spelling difficulty, and the relationship is in the direction expected; that is, more concrete words tended to be easier to read and spell.

The relationship between the frequency of occurrence and reading and spelling difficulty as given in Table I is significant and in the expected direction; that is, more frequent words tended to be easier to read and spell. These relationships are considerably increased when the $\log_2$ of the frequency of occurrence is related to reading and spelling difficulty. This indicated that the relationship of frequency of learning is not linear, but that initial learning responses are of greater importance than later responses.

Table I indicates no relationship between multi-ordinality and concreteness, suggesting that these aspects of meaning are independent of one another. The differential relations of these two variables to reading and spelling difficulty tends to confirm this notion.

Table I indicates a mild, positive relationship between multi-ordinality and frequency of

TABLE I

CORRELATIONS BETWEEN MULTI-ORDINALITY, CONCRETENESS OF SIGNIFICATES, AND THE SPELLING AND READING DIFFICULTY OF WORDS

	Grade Level of 50% Correct Reading	Multi-Ordinality	Concreteness	Frequency of Occurrence	Log_2 Freq. of Occurrence	Word Length
Grade Level of 50% Correct Spelling	.296	-.288	-.196	-.341	-.492	.568
Grade Level of 50% Correct Reading		-.160	-.362	-.193	-.488	.270
Multi-ordinality			-.051	.220	.292	-.251
Concreteness				.053	.274	-.186
Frequency of Occurrence					.537	-.164
Log_2 Frequency of Occurrence						-.393

$r_{.05}$ w/w 147 d.f. = .159
$r_{.01}$ w/w 147 d.f. = .208

occurrence and $\log_2$ of frequency of occurrence. Thus, as the frequency of words increases, there is a tendency for the number of meanings to increase, as was suggested by Zipf (13, 14). It also suggested that the relationship between frequency of occurrence and reading and spelling difficulty may be paramount to the relationship between multi-ordinality and reading and spelling difficulty. A partial correlation, $r_{1.23} = -.208$, indicated the relationship between multi-ordinality and spelling difficulty to be low, barely significant, but at least in part independent of the frequency of occurrence of the word. Words with a larger number of meanings, then, tend to be easier to spell.

Similarly, the correlation between multi-ordinality and reading difficulty with frequency held constant, is $r = .022$, indicating that multi-ordinality relates to reading difficulty only as a function of the relationship it bears to frequency. The number of meanings of a word, then, does not effect the difficulty of that word for the reading learner.

The relationship between concreteness and frequency of occurrence was not significant. However, when $\log_2$ of frequency of occurrence is considered, this relationship is significant. This would indicate that concreteness relates to frequency when the latter is converted to the $\log_2$ scale. In essence, the frequency of concrete words tends to be higher. The relationship of concreteness to reading difficulty and spelling difficulty with $\log_2$ of frequency constant are .271 and -.094, respectively, indicating the concreteness of the meaning relates to the difficulty of learning to read the word. This relation is independent of the frequency of occurrence. The relation of concreteness to spelling difficulty is only a function of the relation of concreteness and frequency.

The three variables may be used to predict spelling difficulty. A multiple correlation between multi-ordinality, concreteness, and frequency of occurrence, with spelling difficulty, is $R = .527$. For reading difficulty multi-ordinality does not make a contribution to the multiple correlation. Thence frequency and concreteness together produce a multiple correlation of .543. The multiple regression equation for predicting spelling difficulty from meaning variables is as follows:

$$SD = 7.22 - .031M - .069C - .219L_2F$$

where:

SD equals spelling difficulty,
M equals multi-ordinality,
C equals concreteness, and
L_2F equals $\log_2$ of frequency of occurrence.

Similarly the multiple regression equation for reading difficulty is:

$$RD = .928 - .531C - .623L_2F$$

where

RD equals reading difficulty,
and the other symbols above apply.

Discussion

Of the three variables, the frequency of occurrence relates most highly to both spelling and reading difficulty. Whether this is a function of the fact that reading materials are graded in relation to frequency is open to question.

Of the other two variables, concreteness is most highly related to reading difficulty, whereas multi-ordinality is most highly related to spelling difficulty. The inference that may be drawn from these data reflect the basic differences in the spelling and reading processes. Reading is essentially a process in transmitting meaning. Hence, the more familiar the significate, in terms of the senses which can be applied to it, the more meaning it will transmit. Spelling, on the other hand, appears to require only an understanding that there is a meaning for the word.

It may be inferred from the data that increased multi-ordinality would increase the probability of an awareness of some meaning of a word. Hence, where a generalized knowledge of the meaning of a word is required for learning, as in spelling, multi-ordinality relates to learning difficulty. The child learning to spell is primarily not concerned with the communication process, but is concerned with the task of learning to spell. Hence, the fact that a word is meaningful, is, for him, sufficient, and the greater the number of meanings for a word, the greater the probability that he will be aware of one of them.

Concreteness of the significate, on the other hand, relates to the specifiability or tangibility of the meaning. This variable then is a determiner of learning difficulty when a specific meaning is required of the learner. Hence it is most applicable to reading learning.

The results imply that reading teaching will be more effective if greater emphasis is placed upon words which have specific and concrete meanings, and also if greater emphasis is given to making these words meaningful to children. The experience program in spelling teaching is a move in this direction. Individualized reading tends to be a second force for the child will select those materials which are meaningful to him, but the data also suggest the organization of reading materials in relation to concrete words and experiences within the child's realm. Certainly this should be done by the teacher. The concreteness of the significate should represent a second criterion in the selection of words for reading teaching materials. An emphasis on concreteness would tend, over the years, to increase the relationship between the concreteness of the significate and the learning difficulty of the reading words.

Frequency of occurrence measures the familiarity with a symbol, and hence its meaning which is general in the population. Hence it is an index of the number of repetitions of the word and relates to all language learning situations. The data indicated that the relation between frequency and learning difficulty is not a linear function, but the greatest increment of learning occurs in the initial trials.

Summary

Multi-ordinality of meaning, concreteness of the significate, and the $\log_2$ of frequency of occurrence of words related to the difficulty of learning to read and spell words. Multi-ordinality related to spelling difficulty, but not to reading difficulty. Concreteness related to reading difficulty, but not spelling difficulty, indicating a difference in the demands of the reading and spelling processes. Log_2 of frequency of occurrence was superior to multi-ordinality, concreteness, and simple frequency of occurrence in predicting learning difficulty.

REFERENCES

1. Gates, A.I., Spelling Difficulties in 3876 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1937.
2. Green, H.A., The New Iowa Spelling Scale. Iowa City: University of Iowa, 1953.
3. Johnson, W.D., People in Quandries. New York: Harper and Brothers. 1946.
4. Korzybski, A., Science and Sanity, third edition, Lakeville, Connecticut: The International Non-Aristotelian Library Publishing Company, 1948.
5. Lorge, I., "Predicting Reading Difficulty of Selections for Children," Elementary English Review, 16:229-33, 1939.
6. Lorge, I., and Thorndike, E.L., Teachers Word Book of 30,000 Words. New York: Teachers College Bureau of Publications, 1940.
7. Marks, M.R. and Ollie, Jack, "Verbal Context and Memory Span for Meaningful Material," American Journal of Psychology, 65:298-300, 1952.
8. Noble, C.E., "The Meaning-Familiarity Relationship," Psychological Review, 60:89-98, 1953.
9. Rinsland, H.D., A Basic Vocabulary of Elementary School Children, New York: The Macmillan Company, 1945.
10. Thorndike, E.L., Teachers Word Book. New York: Teachers College, Columbia University, 1921.
11. Thorndike, E.L., Teachers Word Book of 20,000 Words, New York: Teachers College Bureau of Publications, 1932.
12. Webster's New World Dictionary of the American Language, Joseph H. Friend and David B. Guralnik, general editors. Cleveland and New York: The World Publishing Company, 1954.
13. Zipf, G.K., Psycho-Biology of Language, New York: Houghton Mifflin, 1935.
14. Zipf, G.K., "The Meaning-Frequency Relationship of Words," Journal of General Psychology, 33:251-56, 1945.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

VALUES AND ACHIEVEMENT IN A FRESHMAN PSYCHOLOGY COURSE

WILLIAM D. WINTER*
San Jose State College
San Jose, California

ALTHOUGH MANY years of work have gone into refining intelligence and aptitude tests which would accurately predict school achievement, we seem to have approached a practical limit to their predictive validity. Most good tests of this type generate correlations averaging about .50 with grades (6). Interest has grown in determining the influence of nonintellectual factors such as values and motivation on achievement, and we have seen the recent development of paper-and-pencil personality inventories which predict high school and college grades, often as well as or better than the usual aptitude tests (3, 4, 5).

One of the many factors which have been hypothesized as affecting school performance is the degree of similarity in values between a pupil and his teacher. Sociologists have pointed out, for example, that our school system is dominated by middle class values. The lower class child cannot accept these values, cannot identify with his middle class teacher, and is likely to fail academically, regardless of his intellectual abilities (7). In a related, but somewhat different context, Carter found that achievement in a college course in educational psychology was related to the degree of scholarly motivation possessed by a student and the degree to which he accepts or rejects the school's values of intellectual achievement (2).

The present study is an outgrowth of these findings, and attempts to measure the similarity in values between a group of college students and their instructor, and to relate this similarity to their academic achievement in his class.

Method

The subjects of this study were 34 freshmen men enrolled in a required course in general psychology taught by the same male instructor. This was the instructor's first teaching experience at the college, so his classes were unselected on the basis of his reputation with the student body. In addition, he was a recent graduate of a well-known university with a rigorous experimental program, and he represented a relatively clear-cut value system in which intellectual, research, and scientific values stood high.

A modified form of the Semantic Differential, as described by Osgood, and his co-workers (8, 9), was the instrument used to measure values. This is a reliable test which has yielded good results in studies of the meaning of concepts and the semantic space they occupy. In this present study, the students were asked to rate fifteen concepts according to eight adjective pairs, such as good-bad, sweet-sour, etc. One member of the adjective pair was placed at one end of a seven-point scale, the other member at the opposite end. Then the student indicated the direction of his judgment by rating the concept at some point along the scale, giving it a score of one to seven. Osgood, by means of factor analysis, found three main dimensions of connotative meaning for the concepts he studied: value, strength, and activity (8, 9). However, the present study pertained only to values, so the eight adjective pairs were drawn wholly from Osgood's list of highly inter-correlated value adjectives, so as to provide a precise measure of values. An example of the form of the test is given below:

RESEARCH

good __: __: __: __: __: __: __: bad

The fifteen concepts studied were: athletics, books, cheating, college, easy money, good time, grades, homework, play, professor, research, social activities, studying, tests, and work. The eight adjective pairs were: beautiful-ugly, clean-dirty, fair-unfair, good-bad, kind-cruel, nice-awful, sweet-sour, and valuable-worthless. Since only the value dimension was studied, Osgood's suggested

---

* The author wishes to express his gratitude to Professor James McGaugh and James Asher for their cooperation and efforts on behalf of this project. Professor Asher was instrumental in developing the particular test of values and in planning the research.

scoring of profile similarity was not used. Instead, the total number of points that each concept received on all eight adjective pairs was the basic score indicating the value of that concept for a given individual. Since each adjective pair has a maximum score of seven (which indicates low value), a score of fifty-six is the highest obtainable for any concept.

Each student was asked to fill out the Semantic Differential quickly, according to their first impressions, and were told that there were no right or wrong answers. Previously, their instructor had also filled out the same test. When the students had finished, they were given another copy of the test and told to predict the responses that their instructor had given. They also filled out a brief questionnaire giving their age, parents' education, etc. Their entrance examination scores on the ACE Psychological Examination, and their final ranking in the course, based on test grades, were obtained at the end of the semester. For brevity, these final ranks will be referred to as "grade achievement," since they formed the direct basis for the assignment of letter grades.

For each student, the total value score for any given concept was subtracted from the value score given to it by his professor. The individual discrepancies were summed across concepts to give a measure of overall value discrepancy between student and professor (referred to hereafter as Student-Professor score). This was the basic discrepancy score which was correlated by the rank-order method with grade achievement to test the major prediction that the greater the similarity in values between professor and student, the higher the student's grade. Since the literature points to significant relationships between grades in similar course areas and the ACE - Linguistic and Total scores, these were also predicted (1). The three predictions were tested by one-tail tests of significance, the others by two-tail tests. Father's education, mother's education, and the discrepancy between the professor's values and the student's prediction of his values (Professor-Predicted score) were also explored.

Results

The rank-order coefficients obtained by correlating each of the above variables with grade achievement in elementary psychology are listed in Table I.

It will be seen from this table that our three major hypotheses are confirmed. Grade achievement is significantly related to the ACE - Linguistic and Total scores, and to the degree of similarity in values between student and professor. In addition, grades were also significantly correlated with the student's ability to correctly predict his instructor's values. Father's education and the Quantitative score of the ACE were uncorrelated with grades, while mother's education showed a small but not significant correlation with the criterion.

The two best predictors of class achievement are the Student-Professor discrepancy score and the ACE - Linguistic score. These seem to be measuring disparate factors, as the correlation between them is only -.10. Between Student-Professor and ACE - Total, the correlation is -.05. The value discrepancy and academic aptitude measures appear to be distinct variables, although each correlates with grades.

The correlations with Professor-Predicted are more complex. This score correlates +.46 with Student-Professor discrepancy, -.41 with ACE-Linguistic, and -.39 with ACE - Total. Apparently, the ability to judge the professor correctly is related to intellectual and aptitude factors, as well as to achievement and to similarity between student and professor in values.

Discussion

The major finding of this study is that, for this sample, the more similar in values a student was to his instructor, the higher was his achievement in class. This predicted relationship is congruent with the body of literature in this area and indicates the importance of non-intellectual factors in college achievement. It is probable that since the professor in this case is a scholarly person, a student with discrepant values would be a non-scholar, one who is unmotivated to read and study, although perhaps having the intellectual ability to do so. The correlation between the student's ability to predict the professor's values and his grades indicates that the poor student probably finds it difficult to understand, identify with, and learn from his professor.

When we examine the specific predictions of professor values made by the high and low achievers (based on median class rank) we obtain some insight into the stereotype of the professor held by the poorer students. The low achievers thought their professor would dislike such concepts as: college, good times, play, social activities, and research; and that he would have a higher opinion of cheating and homework. The findings for cheating and research are difficult to interpret. However, the general feeling of the poor students is that the professor is a drudge who likes homework, not a fun-loving person with active social interests. When we examine the professor's actual value scores, we find that for each of the above concepts, the high achievers were better able to guess correctly the professor's true values. These individual concept differences are all in the same direction, and as mentioned previously, the overall score was significantly correlated with grades.

The ACE - Quantitative score and parental education were unrelated to any of the variables studied. Language abilities seem much more important than numerical abilities in predicting grades in this course, which is in accord with previous findings (1). Parental education is probably a less potent variable at the college level than in elementary or high school, since the poor academic records of children with lower class values would keep them

TABLE I

CORRELATIONS BETWEEN VARIABLES AND ACHIEVEMENT IN CLASS

Variable	Correlation
Student-Professor Discrepancy	-.46 **
ACE - T Score	+.36 *
- Q Score	+.13
- L Score	+.40 **
Father's education	+.04
Mother's education	+.26
Professor-Predicted Discrepancy	-.38 *

* Significant at the .05 level, 2-tail test
** Significant at the .01 level, 1-tail test

from being admitted to college even if they desired to attend, which would be unlikely.

It is obvious that a great deal of research remains to be done in this promising area. Replication of this study with different classes and instructors of both sexes is needed in order to determine the generality of these results. It would be interesting to determine if the various departments of instruction or the teaching staff as a whole has a set of common values which students must accept in order to succeed in college. What is offered here is a technique by which these problems could be studied. It is suggested that other instruments be tried out which do not require the many hours of laborious clerical work needed to score the Semantic Differential.

Summary

A modified form of the Semantic Differential was administered to 34 male students enrolled in freshman psychology. This test was designed to measure their value judgments of fifteen concepts. In addition, their instructor took the same test, and the students were asked to predict his values.

The findings of this study are that grades achieved in this class are significantly correlated with:

(a) The similarity in values between professor and student.
(b) The student's ability to predict the professor's values correctly.
(c) The student's scores on the ACE - Linguistic and ACE - Total scales.

Parental education and the ACE - Quantitative score were uncorrelated with grades.

REFERENCES

1. Berdie, R., Dressel, P., and Kelso, P., "Relative Validity of the Q and L Scores of the ACE Psychological Examination," Educational and Psychological Measurement, XI (Winter, 1951), pp. 803-812.
2. Carter, H. D., "What Are Some of the Basic Problems in the Analysis of Study Techniques?" California Journal of Educational Research, II (September, 1951), pp. 170-174.
3. Duffy, Elizabeth, "A Critical Review of Investigations Employing the Allport-Vernon Study of Values and Other Tests of Evaluative Attitude," Psychological Bulletin, XXXVII (October, 1940), pp. 597-612.
4. Gough, H., A Nonintellectual Intelligence Test," Journal of Consulting Psychology, XVII (August, 1953), pp. 242-246.
5. Gough, H., "The Construction of a Personality Scale to Predict Scholastic Achievement," Journal of Applied Psychology, XXXVII (October, 1953) pp. 361-366.
6. Krech, D., and Crutchfield, R., Elements of Psychology, New York: A.A. Knopf, 1958.
7. Lindgren, H. C., Educational Psychology in the Classroom, New York: John Wiley, 1956,
8. Osgood, C. E., and Suci, G. J., "Factor Analysis of Meaning," Journal of Experimental Psychology, XX (November, 1955), pp. 325-339.
9. Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., The Measurement of Meaning, Urbana: University of Illinois Press, 1957.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

A RE-EVALUATION OF THE HEATH RAILWALKING TEST

CORNELIUS P. GOETZINGER*
Kansas University Medical Center
Kansas City, Kansas

MOTOR ABILITY and locomotor coordination are usually regarded as primary aspects of maturation and development. The latter, as evidenced in the age at which a child begins to walk, is frequently taken as supportive confirmation in the diagnosis of mental retardation. That there is not a clear-cut relationship between motor and mental abilities even at early ages, but instead, a large amount of overlapping of functions, has been stressed by Bayley (1) and Gesell and Amatruda (5). Nevertheless, Myklebust (11)(12)(13) has pointed out the significance of a motor evaluation in contributing to a differential diagnosis of aphasia, deafness, autism and mental retardation in children.

Motor tests for clinical use were developed by Bayley (1), Oseretsky (3), Heath (8), and others. Heath, in particular, devised a rail-walking test to measure gross locomotor ability. It would appear from a review of the literature that no subsequent research has been directed toward confirming Heath's original norms for children by utilizing a sample from a different part of the country.

Locomotor ability as exemplified in the act of walking unaided normally develops in the child between the ages of 12 and 15 months. Delay in walking or poor locomotor ability may be contingent upon mental retardation or injury to the central nervous system. With regard to the former, the relationship between mental age and locomotor ability as measured by the Heath Rails Test seems to depend upon the type of mental defect. Thus, Heath (10) reported a partial correlation, with chronological age controlled, of .64 between mental age on the 1916 Stanford Binet Test and the Heath Railwalking Test for 50 endogenous mentally defective children. More recently, Frisina (4) obtained a correlation of .59 (plus-minus .07) between mental age and Heath's with a group of hearing impaired children whom he described as evincing a familial form of mental retardation. When, however, mental defect is associated with accident, disease, etc., there is no relationship between the aforementioned variables. With a population of 80 exogenous mentally defective children Heath (10) found the partial correlation with chronological age held constant between mental age (1916 Stanford-Binet) and railwalking to be .001. This result obviously indicates the absence of a relationship.

In discussion equilibrium as associated with the railwalking test Heath (8) points out that both sensory and motor pathways are involved. The former indicate to the subject that he is off balance. The latter enable him to regain balance. Poor performance per se is often difficult to attribute to either sensory or motor defect, or to central coordinating mechanisms. In addition to the vestibular apparatus consisting of the semi-circular canals, the utricle and the saccule, the cues afforded by vision and the kinesthetic senses likewise contribute to equilibrium. Myklebust (13) has supplied evidence of decreased ability of children in schools for the deaf on the railwalking test, particularly, with reference to the effects on balance of meningitis which destroys the labyrinth. In general, individual differences in railwalking performance according to Heath are probably associated with the central coordinating mechanisms rather than to a disturbance in the transmission of sensory stimuli. It would seem then that the lack of correlation between railwalking ability and mental age in the exogenous defective is possibly due to the variability of the effects of the causative agent upon intelligence, the central coordinating mechanism, and sensory and motor transmission. Conversely, one might reasonably expect a correlation between the two variables in the endogenous defectives who are characterized by a generalized reduction of potential in all areas.

Heath (8) in 1944 proposed an extension of the concepts of endogeny and exogeny to include individuals with normal intelligence. He further suggested that the bulk of the population was made up of endogenous normals who manifested relatively even psychometric profiles, normal railwalking scores, the

---

*Other authors include J. D. Ortiz, Kansas School for the Deaf, Olathe, Kansas
C. L. Rousey, Humbolt College, Arcata, California
D. D. Dirks, Kansas University Medical Center, Kansas City, Kansas

TABLE I

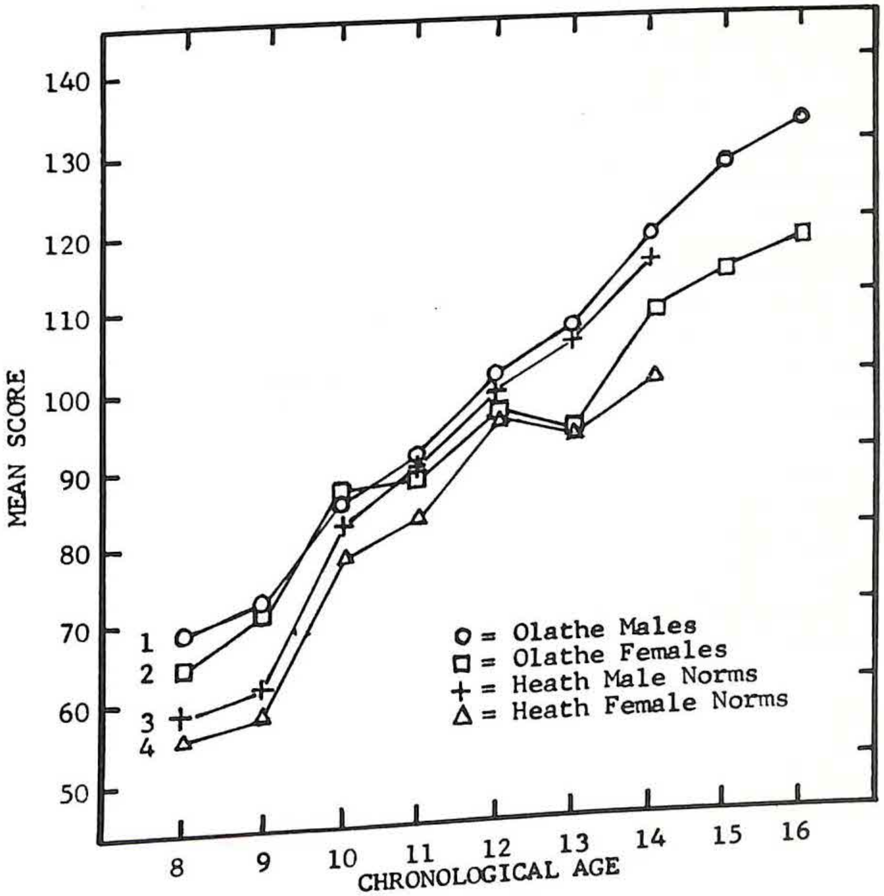
MEANS AND STANDARD DEVIATIONS ON THE HEATH RAIL TEST FOR THE SUBJECTS OF THIS STUDY AND THE HEATH NORMS

C. A. Years	Olathe Subjects				Heath Norms			
	Males		Females		Males		Females	
	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.
6					40.0	18.5	44.4	22.0
7					45.6	21.8	49.0	19.7
8	68.4	29.1	64.2	26.4	59.0	27.8	57.2	24.4
9	71.1	30.6	70.6	31.8	62.2	32.9	59.0	28.8
10	84.1	30.7	86.1	18.7	82.3	31.5	76.7	26.7
11	90.1	16.3	87.5	28.8	88.6	30.2	82.3	33.2
12	101.1	32.0	96.2	18.7	99.0	23.8	95.3	26.4
13	107.4	28.4	93.5	22.1	106.1	32.1	93.0	26.0
14	117.8	24.9	108.1	26.3	115.4	27.5	99.9	32.9
15	127.2	21.2	113.5	18.6				
16	132.4	21.8	117.6	11.8				

Note - There were 10 males and 10 females at each age level or 180 subjects for the age range 8 to 16 inclusive in the Olathe group. The number of sugjects at each age level in the Heath study ranged from 30 to 56.

FIGURE 1

A COMPARISON OF THE MEAN SCORES ON THE RAILWALKING TEST, AT EACH AGE LEVEL, BETWEEN THE HEATH NORMS AND THE OLATHE PUBLIC SCHOOL SUBJECTS



absence of severely debilitating diseases in their case histories, and a familial history not characterized by gross deviations in occupational pattern. Exogenous normal on the other hand, though possessing normal intelligence quotients nevertheless, were prone to have variable psychometric profiles and railwalking scores. In addition, there was a marked incidence of disease in their case histories. On the assumption that the public school population consists predominantly of endogenous normals a further aspect of this investigation pertained to the relationship between mental age and railwalking ability.

The purpose of this study was then: to re-evaluate the Heath railwalking norms by administering the test to a group of children from a different section of the country (the norms were established on public school children in Philadelphia) and second, to explore the relationship between railwalking ability and mental age as measured by the Goodenough Draw-A-Man Test.

Procedure

Subjects One hundred and eighty children from the primary, intermediate, and secondary departments of the Olathe public school system (a small, mid-western town in Kansas) were used as subjects. There were 90 males and 90 females, ranging in age from eight years to sixteen years-eleven months. Ten students of each sex at age levels 8, 9, 10, etc., through age 16 were selected for study.

Testing A battery of measurements which included a pure tone hearing test by air conduction, the Heath Railwalking Test and the Goodenough Draw-A-Man Test were administered. Only children with normal nearing bilaterally were included in the study. Therefore, each subject was first given the pure tone hearing test at frequencies 250, 500, 1000, 2000, and 4000 cps for each ear. Any subject who did not pass the hearing test at a 10 db level or less for the aforementioned frequencies in each ear was dropped from the study. The criterion of normal hearing was rigidly adhered to so as to rule out as completely as possible vestibular dysfunction and its consequent effects on balance. Likewise, no subject with an apparent locomotor disturbance was included.

The instrument used for the hearing test was a Maico F-1 audiometer which had been calibrated at the factory previous to the investigation, and which was checked periodically during the collection of the data on an Allison Artificial Ear Assembly, Model 3A, at the University of Kansas Medical Center Hearing Clinic. If a child passed the hearing test and showed no apparent locomotor problem he was given the Heath Rails and the Draw-A-Man Tests.

Results

Heath Norms and Olathe Data. In Table I are shown the means and standard deviations of the subjects of this study and the normative data of Heath.

It will be observed that no data were collected in this study at ages six and seven, and that the range was extended to include age 16. The reason for the changes was to provide a suitable reference for a subsequent study of deaf children.

To facilitate the comparison of the data Figure I was prepared. The similarity of the curves for both male and female subjects with the respective norms is apparent. Of particular interest is the notch at age 13 in the female curve. It corresponds identically with the Heath norm. This notch may possibly reflect a disturbance in balance associated with female adolescent body-changes.

When the data were analyzed statistically by means of 't' ratios (14) none of the differences at each age level between the means of the males of this study and the Heath male norms, or between the females of this study and the Heath female norms was significant at the five per cent level of confidence.

These findings therefore offer confirmative evidence that the Heath norms are applicable to school children in a mid-western community.

Relationship Between Mental Age and the Heath Rails

Next, rank-order correlations were computed between mental age and the railwalking scores for each sex separately and for the combined sexes at each age level. Of the 27 correlations which were computed only one reached significance at the five per-cent level. It is of course true that the number of cases at each age level was small. Nevertheless, if a relationship exists between the two variables a positive trend would have been expected. As an additional check partial correlations were computed between mental age and railwalking scores with chronological age held constant for 55 males and 56 females whose intelligence quotients ranged from 85 to 115. The obtained partial correlations of .146 for the males and .076 were not significant at the five per-cent level of confidence. Little if any relationship therefore, could be demonstrated between mental age and railwalking ability for these subjects from the public school. Although the criteria of endogenous normals as set forth by Heath were certainly not fully satisfied in this study, nevertheless, in view of the agreement between the Heath Rail Norms and our subjects it seems unlikely that our sample was grossly atypical, being heavily weighted with exogenous normals. The findings therefore suggest that locomotor ability as measured by the Heath Rails Test is a poor index of mental level for the age range 8 to 16 years, and as a consequence, not indicative of cerebral functioning which is associated with mentality.

Summary

The purpose of the study was first: to re-evaluate the norms of the Heath Railwalking Test, and second: to examine the relationship between railwalking ability and mental age as measured by the Draw-A-Man Test. Ninety male and ninety female subjects ranging in age from eight to sixteen years inclusive were given the aforementioned tests if a previous hearing examination had revealed normal acuity for pure tones.

The results indicated first: that there were no statistically significant differences between the Heath Railwalking norms for either boys or girls and the subjects of this study. The finding suggests that the Heath norms are applicable for general use. Second: a significant relationship between locomotor ability and mental age was not found in this study. Locomotor ability therefore, does not appear to be a good index of mental ability as measured with the Draw-A-Man Test even though the subjects could be reasonably regarded as endogenous normals.

REFERENCES

1. Bayley, Nancy. The Development of Motor Abilities During the First Three Years. Washington: Society for Research in Child Development, National Research Council, 1, 1935.
2. Callahan, Fern E. A Study of the Effect of the Loss of Visual Cues and of Flat-Footedness on Locomotor Ability in Relation to Hearing Loss. Unpublished Master's Thesis, University of Kansas, 1958.
3. Doll, E. A. The Oseretsky Test of Motor Proficiency. Minneapolis: Educational Test Bureau, 1946.
4. Frisina, R. Motor Ability. The Health, Physical Education, and Recreation Review, 1, 1957, 2.
5. Gesell, A. L. and Amatruda, C. S. Developmental Diagnosis. New York: Paul B. Hoeber, Inc., 1948.
6. Goodenough, F. L. Measurement of Intelligence by Drawings. New York: World Book Company, 1926.
7. Goodenough, F. L. "A New Approach to the Measurement of the Intelligence of Young Children." Pedagogical Seminary and Journal of Genetic Psychology. 33, 1926, 185-203.
8. Heath, S. R. "Clinical Significance of Motor Defects, with Military Implications." American Journal of Psychology. 57, 1944, 482-499.
9. Heath, S. R. "The Military Use of the Railwalking Test as an Index of Locomotor Coordination." Psychological Bulletin, 40, 1943, 282-284.
10. Heath, S. R. "Rail-Walking Performance as Related to Mental Age and Etiological Type Among the Mentally Retarded." American Journal of Psychology. 55, 1942, 240-247.
11. Myklebust, H. R. Auditory Disorders in Children. New York: Grune and Stratton Company, 1954.
12. Myklebust, H. R. "The Relationship Between Clinical Psychology and Audiology." Journal of Speech and Hearing Disorders. 14, 1949, 98.
13. Myklebust, H. R. "Significance of Etiology in Motor Performances of Deaf Children with Special Reference to Meningitis." American Journal of Psychology. 59, 1946, 249-258.
14. Walker, R. N. and Lev, J. Statistical Inference. New York: Henry Holt and Company, 1953.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

STUDENT MORTALITY AND RELATED FACTORS

NORMAN GEKOSKI and SOLOMON SCHWARTZ
Temple University
Philadelphia, Pennsylvania

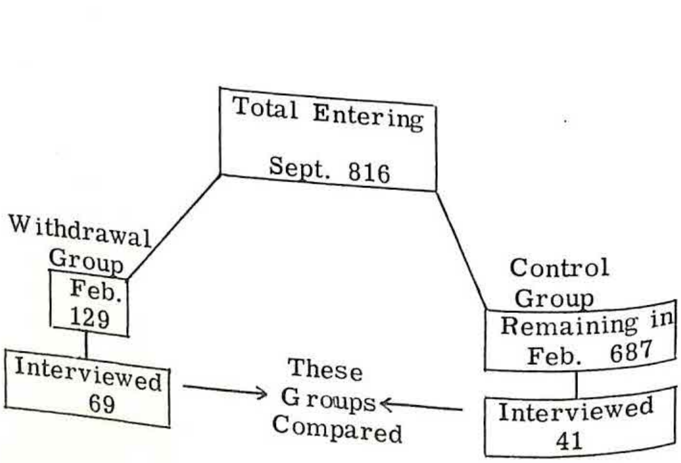
THE EMPHASIS on science education since the launching of Sputnick I has tended to obscure other important problems in American education. One of these is the problem of student mortality -- or "Why Students Leave School". This problem exists at the secondary and the collegiate levels. We are concerned here with the latter. It has been estimated that approximately 50% of the students who enter college leave before they receive a degree, and most of these leave in the first year. This situation presents serious problems to the student in terms of blighted hopes and frustrated ambitions; to the colleges in terms of useful services not rendered; and to the nation in terms of wasted human resources.

The psychology department of Temple University recently undertook a study of this problem. The study was designed to find in what respects students leaving college differed from those who remained to complete their work. Previous investigations had not concentrated upon this aspect of the problem. An analysis of the literature on the subject revealed several weaknesses in previous approaches. In the first place most of the studies concentrated upon factors related to the achievement of those who remained in school, rather than upon those factors associated with withdrawal. Secondly, when they had examined factors which contributed to withdrawal they concentrated on one or two causes, rather than investigating several causes. Finally, when describing the withdrawal groups they have listed its characteristics without comparison to those who remain in school. Such description without comparison does not shed light on the real factors contributing to student mortality. For example, if one hundred students have financial difficulties and fifty leave school while the others remain, then factors other than financial played the decisive role. This study has emphasized a comparison of withdrawing students with those who remain in school.

A recent entering class at Temple University was chosen as the group to be studied. This class consisted of 816 freshmen registered in the Liberal Arts College, the Teachers College, and the School of Business and Public Administration. Of this group 262 or 32% did not register the following September. One-half year later an additional 8% dropped out. This represents a 40% loss in one and a half years with 2/3 of the loss coming in the first year. This emphasizes the crucial nature of the first year.

In conducting this study complete data were collected on all of the students. This included entrance test information, application blank information and information obtained in a personal interview. This report concerns itself primarily with an analysis of the information obtained from the interview.

To point out the characteristics in which withdrawing students differ from remaining students a "control group" was used. Employing sampling techniques, students were selected from the remaining college population and were used as a standard against which the withdrawal group was compared. It proved impossible to interview all of the students. Therefore, systematically selected samples were obtained from each group, and these people were interviewed. The following chart presents the nature of the groups under consideration and those actually interviewed. It is to be noted that this report deals with comparisons between those who left the University after the first semester, and those who registered for the second semester.



A careful analysis of the characteristics of the sample groups showed a striking similarity between the samples and the groups from which they came. The investigators are confident that the withdrawal interview group is representative of the total withdrawal group, and that the control interview group is representative of the total remaining group.

The students in both groups were tested for mental ability, occupational interest, personality, and study habits. In addition each student's background was investigated. Factors such as age, sex, type of high school attended, marital status, etc., were recorded for each subject in the study. Finally, each student was interviewed. The interview questionnaire revealed such information as reasons for withdrawal, vocational goals, types of extra-curricular activities and opinions concerning courses, instructors, university administrators, and facilities.

The study has yielded much significant information to date. In comparing the test scores of the two groups it was found that there was a significant statistical difference between them on the scholastic aptitude test, the reading test and the social adjustment section of the personality test. In all of these cases the withdrawal groups scored lower. This lower level of scholastic ability no doubt contributed to their withdrawal. However, the ability of the withdrawals was still sufficiently high that they could have been expected to succeed in college under optimum conditions.

When questioned concerning the chief reason for withdrawing from college the withdrawal group responded as follows:

Change of interest or plans	21%
Dissatisfaction with courses and/or university	16%
Financial difficulties	12%
College adjustment problems	12%
Job interference - to go to work	12%
Health	9%
Low Grades	9%
Entering armed forces	5%
Commuting problems or others	4%

However, interviews with the control groups indicated that they too had similar difficulties, yet they remained in school.

Further analysis of the interview questionnaire indicated that there were significant respects in which the two groups differed. Some of these were as follows:

1. Twice the percentage of the withdrawal group as compared with the control group felt that their courses were preparing them poorly or very poorly for their vocational objectives. Conversely, twice the percentage of the control groups as compared with the withdrawal groups, felt that their courses were preparing them well or very well for their vocational objectives. These facts would indicate that one's attitude toward the quality of vocational training is an important factor in determining whether he leaves or remains in school. This finding is particularly significant when it is noted that 36% of this class indicated that their expectations from college were primarily vocationally oriented.
2. In an overall evaluation of the faculty considering teaching ability, interest, knowledge, organizational ability, etc., the withdrawal groups rated them much lower than the control group. Twenty-one per cent of the withdrawal groups as compared with four per cent of the control group rated the faculty poor or very poor. On the other hand 64% of the control groups as compared with 43% of the withdrawal group rated the faculty good or very good.
3. There were significant differences between the two groups in their attitudes toward their faculty advisors. Fifty-two per cent of the withdrawal students could not name their advisors. This was true of only 22% of the control group. In addition only 5% of the withdrawal students felt that their advisors had had a favorable effect on their progress in school. On the other hand 41% of the control group felt that their advisors had had a favorable effect on their progress. Finally, a significantly greater percentage (63%) of remaining students had had personal contact with their advisors than had withdrawing students (45%).
4. Remaining students were much more aware of campus agencies which might help with student problems than were the withdrawing students. 44% of remaining students had received such assistance while this was true only of 23% of the withdrawing students.
5. Concerning knowledge of degree requirements, a significantly greater percentage (83%) of remaining students than withdrawals (61%) felt that they had adequate knowledge on this subject.
6. An evaluation of the attitude toward extra-curricular activities indicates certain significant differences between the two groups. Sixty-three per cent of the remaining students rated such activities as good or excellent. Only 42% of the withdrawing students gave these activities so high a rating. In addition the remaining students participated in extra-curricular activities to a greater degree than did the withdrawing students. Finally, the remaining students had a more favorable opinion of the administration's control of extra-

curricular activities than did the withdrawing students.

7. In considering the effect of outside conditions such as home life, social life, religious life, etc., upon progress in school a significantly larger percentage of the remaining students considered such conditions to have had a favorable effect. This was true of 36% of the control group. It was true of only 8% of the withdrawal group.
8. Finally, the remaining students felt that their college experiences have fulfilled their expectations upon enrolling to a greater degree than do the withdrawing students.

The final question in the interview was asked of the withdrawing students only. It asked what the university might have done to prevent the student's leaving school. The responses were as follows:

Nothing	41%
More guidance-counseling	27%
Improved courses and instruction	11%
More personal attitude toward student	5%
Financial aid (scholarship, job, etc.)	5%
Other and "Don't Know"	11%

It is significant that approximately 50% of the withdrawing students felt that some action on the part of the university might have been instrumental in keeping them in school.

The results which this study has yielded to date point to certain courses of action which colleges and universities might take to curtail the rate of student withdrawal. The investigators made certain recommendations to the administration of the University. Among these the following were most important:

1. A re-examination of the curriculum, its philosophy and its content.
2. Making instruction more effective. This would require the selection of teachers on the basis of their ability to present material, not only on their academic backgrounds and achievements. It might also involve in-service training and closer supervision on the part of department heads. Attention should be given also to having freshmen taught by the best instructors and in smaller classes.
3. A greatly expanded and well-publicized counseling service with emphasis on early discovery and identification of student problems. This requires that the whole staff be "guidance-minded" and alert to the type of problems which interfere with the student's success in school.
4. The development of a program for improving the relationships between the student and faculty, particularly the student's advisor. This might include regularly scheduled compulsory interviews.
5. An improved orientation program for new students which would, among other things, make them more aware of campus facilities and agencies which exist to help the students with their problems.
6. There should be a closer relationship between the university and surrounding high schools for the purpose of achieving better continuity in curricula.
7. The stimulation of a greater interest in, and awareness of, withdrawal symptoms on the part of all university agencies and personnel. Early identification of these symptoms with appropriate action taken might save many students who otherwise would leave school.
8. The establishment of a routine system for obtaining an exit interview from students before they leave or as soon as possible thereafter. A system for interviewing a student before he actually leaves has many practical advantages. First, it will obtain information on his reasons for leaving which is accurate, pertinent, and fresh. Secondly, it will have a favorable effect on student morale in general, since it will indicate administrative interest and efficiency. Finally, it might reveal reasons for leaving that can be eliminated and thus forestall the contemplated withdrawal.
9. That a committee be established to deal with the problem on a permanent basis. Such a committee should undertake the implementation of the foregoing recommendations.

In addition to the foregoing recommendations the investigators have emphasized to the University administration the need for continued research of this type. Much additional information is required if this problem is to be dealt with more successfully in the future than it has been in the past. Certainly, a more successful attack upon this problem is necessary if colleges and universities are to achieve their primary function of the optimum development of the human resources and latent talent of American youth.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

A COMPARISON OF THE WISC PATTERNS OF MALE RETARDED AND NON-RETARDED READERS

DONALD NEVILLE
George Peabody College for Teachers, Nashville, Tennessee

THE PROBLEM of whether retarded readers perform in a typical way on the WISC has received some attention in previous research. The results have not been conclusive although some agreement is evident. Results indicating that the verbal scores of retarded readers are significantly lower than their performance scores are common to most studies. However, the reported patterns of subtest scores differ considerably. Graham and Kamono (5), using the WAIS, found unsuccessful readers significantly low in all verbal tests and the digit symbol subtests (same as coding on WISC). Other authors (1, 2, 4, 7) agree that retarded readers tend to score significantly low on the arithmetic and coding subtests although Burks and Bruce (2) and Altus (1) found the information subtest score to be low also. Burks and Bruce concluded further that retarded readers scored significantly high on the comprehension, picture arrangement, and block design subtests. The methods employed by these several researchers varied greatly and the experimental groups were often exceedingly small, less than fifteen.

The Problem

The objective of the present study is to compare the subtest patterns of the Wechsler Intelligence Scale for Children of male retarded readers having IQ's of 90 or above, with male non-retarded readers, also of average or higher IQ's. It is hoped that the results may add to our understanding of the types of intellectual tasks in which retarded readers excell as well as those which they find difficult.

Procedure

In order to accomplish the stated purpose, it was decided to compare the WISC performance of two matched groups. First, 53 retarded readers, children with total WISC IQ's of 90 or above and reading two or more years below grade level, were selected. These subjects represented the total population meeting these criteria and referred to the University of Florida Reading Laboratory and Clinic in the years 1957 and 1958. The Florida Reading Scales, an individually administered test prepared locally, was used to estimate reading grade level. An attempt to match by pairs, this group with a non-retarded group, reading at or above grade level, was then made. This matching attempt resulted in being able to match 35 of the original 53 retarded readers on the following factors: 1) total WISC IQ, within one point, 2) grade level, and 3) sex, all were boys. In addition the subjects in both groups had been referred to the clinic because of a suspected reading problem. There were no subjects included in either group who were referred for further physical or psychological evaluation.

The WISC subtest scores were treated by means of a $\underline{t}$ test for the significance of differences between correlated mean differences in accordance with the suggestion of McNemar (6).

Since the digit span subtest is considered by Wechsler (8) to be supplementary, it was not administered to all subjects. The data for this subtest is based on the 22 pairs for which it was available.

Results

As can be seen from Table I, the scores of retarded readers differed significantly from non-retarded readers in performance and verbal tasks in general as well as in the information, arithmetic, digit span, picture arrangement, and block design subtests. In the first three the retarded reader group was lower and in the latter two higher. All of the above differences are beyond the .01 level of confidence.

Discussion

According to Cohen (3), a factorial analysis of the WISC yields five relatively separate factors. He identifies these factors and the subtests of which they are comprised as follows: Factor A: Verbal

TABLE I

COMPARISON OF WISC SCORES OF RETARDED AND NON-RETARDED READERS

WISC Score	Retarded Readers	Non-Retarded Readers	Mean Difference	Standard Error Mean Difference	t Ratio
Total IQ	100.3	100.5	-.17	.16	1.0625
Verbal IQ	95.3	101.4	-6.11	1.05	5.8190*
Perf. IQ	105.9	99.2	6.57	1.64	4.0061**
Info.	8.7	10.0	-1.31	.40	3.2750*
Comp.	9.7	10.7	-.94	.49	1.9183
Arith.	8.9	10.8	-1.46	.51	2.8627*
Simil.	8.9	9.9	-.09	.45	.2000
Vocab.	9.7	10.2	-.60	.53	1.1321
Digit Sp. [1]	7.9	9.6	-1.77	.54	3.2777*
Pict. Comp.	11.3	10.5	.71	.64	1.1094
Pict. Arr.	11.9	10.0	1.69	.53	3.1887**
Bl. Design	11.3	9.9	1.46	.50	2.9200**
Obj. Assign.	11.0	10.1	.83	.51	1.6275
Coding	8.9	9.0	-.11	.46	.3391

* Retarded readers significantly low, P = .01.
** Retarded readers significantly high, P = .01.
[1] For this subtest N = 22, all others N = 35.

Comprehension I (information, arithmetic, and similarities), Factor B: Perceptual Organization (block design and object assembly), Factor C: Freedom from Distractability (digit span), Factor D: Verbal Comprehension II (comprehension and picture completion), and Factor E: Quasi Specific (coding and picture arrangement). Cohen further notes that Factors A and D are interrelated but that the main difference seems to be the degree to which they depend on the acquisition of formally learned material. Factor A is largely a sample of "school type" learning while Factor D depends more on the application of incidental learning although both are dependent on verbal communication.

Applying these findings we note that retarded readers tend to lack abilities in Factors A and C. At the same time they do well on one subtest in both Factors B and E. Of the three subtests comprising Factor A, "similarities" appears least related to formal learning. It seems reasonable to conclude, then, that retarded readers do poorest in those subtests most nearly resembling school taska and those requiring concentrated attention. Their best performances do not show any logical pattern, factorially. We would point out, however, that they occur in those subtests clearly removed from verbal skills and in activities divorced from school tasks. Further, one is tempted to hypothesize that their excellent performance on the picture arrangement subtest results from long practice at using pictures as clues to the context of the printed page which they are unable to read.

These results raise the question as to whether poor readers are inherently less endowed in some of these areas or whether they fail to develop in them because of their lack of reading ability. From the pattern of results the author tends to accept the latter explanation but further research is needed to confirm this belief.

These results may have some implications for working with retarded readers in a remedial setting. Since they appear to have certain relatively strong and other equally weak functioning areas, it would seem logical to attempt to capitalize on the strengths while minimizing the weaknesses. As a group it is believed that they would profit most from a relatively non-verbal approach utilizing mostly kinesthetic and visual methods of instruction. It might be feasible in working with younger children to attempt to improve their weaker areas through specific types of activities, but the author is inclined to believe this approach to have little hope of success when working with children beyond the primary grades.

Summary

The problem was to attempt to ascertain whether retarded readers exhibited a pattern of WISC subtest scores which was significantly different from a matched group of non-retarded readers. The groups, containing 35 pairs, were matched in total IQ, sex, and grade level.

The results indicated that retarded readers did score significantly low in information, arithmetic, and digit span. They scored significantly high in picture arrangement and block design. It was noted that their low scores seemed to be related to scholastic types of tasks and limited ability to concentrate while the high scores were somewhat removed from formal types of learning. Possible teaching implications of the differential performances were mentioned.

REFERENCES

1. Altus, Grade. "A WISC Profile for Retarded Readers," Journal of Consulting Psychology, XX (1956), pp. 155-156.
2. Burks, H. F., and Bruce, Paul. "The Characteristics of Poor and Good Readers as Disclosed by the Wechsler Intelligence Scale for Children," Journal of Educational Psychology, XLIV (1955), pp. 488-493.
3. Cohen, Jacob. "The Factorial Structure of the WISC at Ages 7-6, 10-6, and 13-5," Journal of Consulting Psychology, XXIII (1959), pp. 285-299.
4. Graham, Earl E. "Wechsler-Bellevue and WISC Scattergrams of Unsuccessful Readers," Journal of Consulting Psychology, XVI (1952), pp. 268-271.
5. Graham, E. E., and Kamono, D. "Reading Failure as a Factor in the WAIS Subtest Patterns of Youthful Offenders," Journal of Clinical Psychology, XIV (1958), pp. 302-305.
6. McNemar, Quinn. Psychological Statistics (New York: John Wiley and Sons, Inc. 1949), pp. 225-226.
7. Spache, George D. "Intellectual and Personality Characteristics of Retarded Readers," Psychological Newsletter, IX (1957), pp. 9-12.
8. Wechsler, David. A Manual for the Wechsler Intelligence Scale for Children (New York: Psychological Corporation, 1949), pp. 115.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 5, January 1961)

A STUDY OF ACADEMIC ACHIEVEMENT AND PERFORMANCE ON A TEST OF LIBRARY UNDERSTANDINGS

WILLIAM D. JOYCE
Director, Regional Public Library
Levittown, Pennsylvania

Preface

THIS STUDY of the relationship between performance by a graduating class of a teachers college in a test of library understandings and their average grades over the four-year course has been conducted in order to probe the connection between skill in using the college library and scholastic success. The study was originally prompted by concern over the observed lack of understanding in using the library displayed by many students and the knowledge that these same students would be called upon to teach library skills in the schools to which they would go following their graduation.

In order to measure the degree of proficiency of each graduating senior in the use of the library, a test was administered. It was felt that such an examination would reveal the weaknesses (and strengths) of the program in library instruction, most of which took place in the freshman year. The results were reported to the administration and faculty, and the testing program was extended to the next entering class. The deficiencies in learning revealed in the original test resulted in the introduction of library experience units into the junior class program. While the curriculum development brought about by this exploration of library understandings was the original purpose of the study, it raised new questions. One of the most interesting and important of these is "Does a relationship exist between the scores made by students on the library test and their academic achievement as measured by their rank standings at the end of their four years in college?"

It was therefore decided to compare the four-year grade averages of the graduating seniors with their rank standings in the library test.

Population Tested

The group studied consisted of eighty-one seniors graduating from the State Teachers College at Lowell, Massachusetts, in 1957. Fourteen of the eighty-one were in the specialized Music Education Curriculum. Of the remaining sixty-seven in the Elementary Education Curriculum, three were absent when the library test was given.

The fourteen members of the Music Education Curriculum have not been included in the comparison of scores because of the need to keep the tested group homogenous. Music students at this college tend to have classroom and library experiences quite different from that of the Elementary Education students.

The test was administered during the final examination period. The sixty-four Elementary Education students who took the library test will make up the population for this study.

Test Instrument

The test used was the "Library Orientation Test for College Freshmen" by Edith M. Feagley Dorothy W. Curtiss, Mary V. Gaver, and Esther Greene, 1955 edition, published by Teachers College, Columbia University. The test is in nine parts covering the following areas:

I. Definition of terms
II. Interpretation of information on a catalog card
III. Choice of subject headings in the card catalog
IV. Arrangement of headings in the card catalog
V. Literature reference books
VI. Sources of biographical information
VII. Choices of indexes
VIII. Interpretation of information in periodical indexes
IX. Abbreviations commonly used in reference books

The test consists of eighty items of the multiple choice and matching varieties. All but four of the reference sources named in the items were available in the college library at the time the test was given.

Although the test title names college freshmen as the intended subjects, the manual indicates that any undergraduate group may be tested with the instrument.

Results of the Test

A range of thirty-five to seventy correct responses was made by the group. The median score was Fifty-three.

The weakest areas in terms of fewest correct responses were:

A. Literature reference books
B. Abbreviations commonly used in reference books
C. Sources of Biographical information

High scorers tended not to know the items to which Current Biography, Biography Index, and "viz" were the correct answers, while low scorers most frequently missed the answers to questions concerning the abbreviations "sic" and "q. v.," and the alphabetical arrangement of headings in the card catalog.

The individual raw scores were arranged by stanines using Dr. Walter N. Durost's table of scores on a nine-point normalized standard score scale with a median equal to five and a standard deviation of two.

In making up the final raw scores for division into stanines, certain test items were found to be non-discriminating between high and low scorers. Eleven such items were dropped in calculating the final raw scores used.

Comparison of Test Results With Academic Rank

Since academic success for the purpose of this study has been defined as the achieved four-year rank-standings based on course grades, these were obtained from the college administration and translated into stanines. The four-year averages ranged from 4.093 to 2.439.

Results

The coefficient of correlation between academic rank and performance on the library test for this sample is .41. The standard error for a sample of this size is .1259 giving a spread for this coefficient of correlation from .28 to .54. Confidence in the belief that a significant relationship between knowledge of library tools and techniques and scholastic success is further increased by a t ratio of 3.5181 giving a probability P of less than .001 that this coefficient of correlation was not established by chance alone.

Summary

In a comparison of scores made in a test of library skills and understandings by a sample of sixty-four teachers college seniors with their academic rank, a significant and positive relationship was established. Those scoring higher than the median on the library test tended to have higher four-year course averages than those who scored below.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 FEBRUARY, 1961 NUMBER 6

CONTENTS

Summary of Investigations Relating to Reading July 1, 1959 to June 30, 1960—Helen M. Robinson........................ 203

Children's Likes and Dislikes About What They Read—J. H. Coleman and Ann Jungeblut................................ 221

The Relationships Between First Grade Reading Readiness and Second Grade Reading Achievement—Marvin Powell and Kenneth M. Parsley, Jr. 229

Personality Traits in Reading Ability of Western Australian University Freshmen—A. W. Anderson................ 234

Reading Improvement in Military, Government and Business Agencies—Ralph S. Acker............................ 238

⁂

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 6, February 1961)

SUMMARY OF INVESTIGATIONS RELATING TO READING JULY 1, 1959 TO JUNE 30, 1960

HELEN M. ROBINSON*
The University of Chicago

THE NUMBER OF studies of reading published this year was somewhat fewer than during the last two years. The policy of including reports involving either a new or distinctive approach to a problem or the identification of significant issues, which was adopted last year, has been followed in the summary this year.

A clear analysis of exploratory and demonstration research in the field of reading was made by Bloom (7). Although this analysis applies particularly to patterns of grouping, it should be of particular value to teachers or others who wish to modify practice on the basis of research reports.

Summaries Relating to Specific Aspects of Reading

As the number of reports of research increases, so does the number of summaries dealing with particular problems. Sixteen summaries this year deal with a range of topics and cover the age range from reading readiness through college and adult levels. These summaries should be examined carefully by those concerned with the specific areas covered.

"Review of Recent Research on College and Adult Reading" by Bleismer (6). The topics discussed are "reading programs," "reading and study skills and habits," "factors influencing skills and habits," "tests," "eye-movements and controlled reading."

"Some Neural and Glandular Bases of Learning," by Eames (20). Although it deals with learning in general, specific references to reading are frequently made. In addition to neural and glandular bases, a third section reviews "recent studies with learning implications".

"Evaluations of Study-Skills Courses: A Review" by Entwisle (22). The topics include "kind of student," "criterion used," "baseline used," "methods used to correct for bias," and "performance in excess of baseline."

"Summary of Reading Investigations, July 1, 1958 to June 30, 1959," by Gray (25).

"Survey of the Research Related to the Reading Ability of the Gifted," by Sister Josephine (33).

"Research in Reading," by Karlin (34). Topics include "reading readiness," "phonics," "interests," "comics and television," "content fields," "individualized reading," and "machines."

"Research on Problems in Reading Science," by Lockwood (40). Organized under "development of scientific vocabulary," "the level of reading difficulty of textbooks for science," and "the ability of teachers to estimate the reading difficulty of science textbooks."

"Implications of Research on Children's Concepts," by McCullough (41). Includes "what a concept is," "how concepts are formed," "factors influencing concept development," and "what teachers can do."

"The Cloze Procedure—Its Validity and Utility," by Rankin (63).

"High School Reading—1958," by Robinson and Dramer (67).

"Language Arts Research, 1958," by Staiger (79).

"Readiness to Read," by Standish (80). Organized under "the tasks of the teacher of beginners," "ascertaining stages of reading development," "use of reading readiness tests for diagnostic purposes," "the value of reading readiness tests," and "teachers' attitudes towards and recognition of reading readiness."

"Print for Children's Textbooks," by Tinker (87). Discusses "size of type," "length of line," "leading, or space between lines," "styles of type face," "quality and color of paper," and "color of paper and ink."

"Another Look at Reading Interests," by Townsend (89).

"Recent Findings and Trends in Research in Reading," by Traxler (90). Based on (92).

Research in Reading During Another Four Years: Summary and Bibliography, by Traxler and Jungeblut (92).

"The Perceptual Process in Reading," by Vernon (93).

---

* Most of the titles in the bibliography were identified and annotated by Dr. William S. Gray prior to his untimely death in September, 1960. The summary was prepared by the writer.

"Individualized Reading—A Summary and Evaluation," by Witty assisted by Coomer and Sizemore (97).

The Sociology of Reading

The number of studies concerned with the social role of reading, of mass media, and of related factors continues to be relatively small although significant.

Analysis of newspapers. One of the persistent problems in studying the content of written materials is categorizing the content. Bush (10) developed a plan for analysis of news content. Rejecting *a priori* categories limited to the number of rows on a machine punch card, he chose to use the range of elements of reader interest in news stories. Furthermore, he based the categories on logical and psychological dimensions rather than on geographical location. Selecting 47 separate categories which were mutually exclusive, his graduate students proceeded to validate the categories with the help of 18 judges. Based on this preliminary work, Deutschmann coded the 1955 front-page news of 27 western newspapers. Ten of the 40 categories attained statistical significance in interest for either men or women.

Although editorial items are usually selected subjectively by an astute editor, Haskins (28) developed and validated a plan for pre-testing reader interest in them. A thermometer interest scale was developed and accompanied by the title, subtitle and author's name of the proposed news item. Both the title and the thermometer were submitted to samples of potential readers, and later the same persons were interviewed to learn whether they responded in the way that the thermometer rating predicted. Predictions of absolute readership (scores above 90 percent on the thermometer) were in error only about seven percent per item. Total predictions yielded a median coefficient of correlation of .57 with subsequent readership. Further refinement of the technique offers promise for exploration of various forms of mass communication and in the study of individual interests in relation to different factors.

With the advent of the Presidential News Conference and an apparent shift in interest from the legislative branch of government to the White House, Cornwell (12) made an analysis of the *New York Times* and the *Providence Journal* to determine whether items reflected such a shift of interest. Sampling the front page only and measuring the number of inches of space devoted to the President, to both, and to the Congress for nearly 70 years, Cornwell concluded that "the data presented indicate that Presidential news, and by inference, the public's image of the Presidency and its relative governmental importance, has increased markedly and more or less steadily in this century, particularly since World War I."

Hage (26) examined selected items in newspapers considered to appeal to educated persons ("quality") and those designed for the masses in 1928 and 1952. He was concerned with the number of items (editorials, letters, etc.) and mention (a word to a paragraph) made in the body of an article. The traits of the candidates for president were classified as "intellectual," "unreflective" (concerning physical action, knowledge gained through experience and intuitively), "integrity," "morality," "democracy," and "patriotism." Furthermore, these traits were classified as "favorable" or "unfavorable" depending on the content. The references to candidates Adams and Jackson, Stevenson and Eisenhower, were rated for the purpose of comparisons. The data as analyzed did not support the conclusion that anti-intellectualism was more predominant in 1928 than in 1952. However, the newspaper comments about Stevenson's intellectualism were predominantly favorable while those for Adams were generally unfavorable. In contrast, unreflective traits of Eisenhower were highly approved while those of Jackson were more often derogatory. As might be expected, unfavorable comments appeared more frequently in newspapers considered excessively partisan.

The last study dealing with analysis of newspapers is unique in that it focuses primarily on the development of children's newspaper reading. Based on samples of 200 to 400 pupils in grades I through VI and at eighth, tenth and twelfth grades, Schramm and Parker (75) found that the newspaper is the last of the "great media" to which the child is introduced. For example, by age three, TV viewing is done with some regularity by about one-third, while more than a half are listening to books read to them; by school age, a third are looking through magazines, 40 percent are listening to radio and over 80 percent are viewing TV; and finally by age seven, when 73 percent are already reading books, a substantial number begin to read the newspaper. As early as four years, the comics in the newspaper are read to children so it is not surprising that comics are the children's first choice in newspapers. Reading the comics continues to be first choice until Grade XII when local news supercedes them. It is at Grades VII and VIII that regularity of newspaper reading is being established and with it comes the beginning of rounded readership. By Grade XII, 90 percent of both boys and girls report that they read the newspaper several days, or almost every day, per week. Bright students begin earlier with most reading patterns, but between Grades VIII and XII the percentage of newspaper reading of dull students increases markedly. The investigators also explored the rank order "prestige" of responses to the mass media.

Preferred social behavior related to televiewing and reading. Perrodin (57) chose the Behavior Preference Record and a personal data sheet to assess the social values of 352 children in fourth grade. The findings were based on information supplied by students on a personal data sheet. The report deals with only those ranked below average so no contrast can be made with those ranked above average. This part of the study revealed that a large portion of the children's out-of-school hours is spent in watching television. Also those who viewed the least preferred non-democratic types of behavior. Those who read the fewest books preferred "non-cooperative, non-friendly and non-responsible behavior." Comic books were popular with these pupils but those who read the fewest preferred non-democratic behavior. Children who were least successful on the reading achievement tests preferred "non-desirable behavior traits." The children who chose "non-leadership" traits were unique in televiewing and reading habits.

Paperback books. Two studies dealt with the impact of the most recent of publishing media. The first, Blum(9) sought to determine the source and content of 4,868 titles in print during 1955. She found that a large proportion were reprints of hardback books. Of this group, 98 percent came from 165 commercial presses. The majority of both reprints and originals were from established authors. Light fiction was the most popular literary form while serious non-fiction represented a small but important portion. The trend was for the better quality titles to be higher in price and less available to the public than the lower quality.

Among available paperbacks, Zamchick (101) recorded the choices of four eleventh-grade classes purchasing books from the Teen Age Book Club. The classes represented superior and average readers as well as two remedial groups. These 71 students obtained 786 books during the year. Nearly half of the purchases were made by the superior readers. However, one remedial group of 12 students purchased as many titles per student as the superior group. In all except the lowest remedial group, the girls tended to be in the upper quartile in number of books purchased while the boys tended to occupy the lower quartile. Only one title appeared among the most popular in three groups. The investigator cautions, however, that many other books were read and that the report does not accurately reflect the taste of individuals.

Library circulation of "good" books. Goldhor (24) examined the circulation of 291,502 volumes in the Evansville, Indiana Public Library at the end of 1958. He selected 741 copies of 317 titles of adult hard-cover books in a particular range of the Dewey Decimal numbers to classify as: A) title with three or more favorable reviews; B) one or two favorable reviews; and C) no reviews or none favorable. This technique resulted in the classification of 38 percent in the A-value category, 38 percent in the B group, and 24 percent in the C group. A careful examination of circulation records led him to conclude that "the best books are not necessarily the most widely read. If librarians buy good books and poor books in equal quantities, they will be read in approximately equal numbers."

Teacher's professional reading. Two investigations were concerned with the reading habits of teachers. Schubert (76) used a six-item questionnaire with 132 experienced teachers (98 elementary and 34 secondary) in his classes to determine the number of books and journals owned which dealt with teaching reading. Elementary teachers reported owning an average of 2.3 and secondary teachers .91 professional books. Elementary teachers subscribed to an average of 1.6 and secondary to 1.2 journals which include articles about teaching reading. Although it is obvious that he had a very select sample, he concluded that the secondary teachers exhibited a greater appreciation for professional books pertaining to reading than did elementary teachers.

Fisher (23) using a case-study interview, examined the professional reading patterns of 50 teachers in four elementary schools. She found that broad general reading of several types tended to be associated with the quantity of professional reading done, but this was not consistently true. Other factors such as being experienced, securing a degree, the accessibility of materials and participation as a member of a curriculum committee appeared to combine with interest, resulted in wide use of professional materials.

Books Germans are reading about America. A unique analysis of 20 non-fiction books about America was compiled by Lewis (39) and the books were ranked in order of popularity. Then content analysis led to rating each book as portraying positive, neutral, or negative attitudes. The number of copies sold in four types of bookstores showed that 11 were positive, seven neutral and two negative toward America. Shifting interest based on recency of publication produced a popularity rating of 144 for the positive books, 99 for the neutral books, and 32 for the negative books. Thus he suggested that the greatest impact was made by positive books.

The Physiology and Psychology of Reading

Only one study could be classified in the area of the physiology of reading while a number dealt with various aspects of the psychology of reading.

Visual fields and reading achievement. Ong, Schneider and Moray (54) measured the perimetric visual fields of 155 pupils in Grades V and VI, then related these data to reading and intelligence test scores for selected portions of the group. The visual fields for boys were about two degrees larger than for girls and there was no significant difference between the sizes of visual fields of the right and left eyes. Furthermore, width of both horizontal and vertical visual fields were unrelated to power or speed of reading, or to I. Q.

Sex differences in achievement. Clark (11) recorded data for boys and girls separately including chronological age per grade, mental age at chronological ages of 66, 72, and 78 months, and then determined the overlapping distribution of scores. In addition, he used statistical control of life age and mental age when he compared reading achievement of boys and girls at Grades III, V, and VIII. His random samples of 150 pupils at each grade were taken from representative numbers and locations. He found no sex differences in general intelligence or basic skills in reading and arithmetic. However, girls achieved higher than boys in the mechanics of English and spelling tests. His data do not support the proposed practice of differentiating school entrance age for boys and girls. Moreover, the wide range in mental ability and achievement among both boys and girls emphasizes the necessity of dealing with individual differences regardless of sex.

Prediction of scholastic achievement for high school seniors was studied for 36 boys and 33 girls separately by Stinson and Morrison (81). The Differential Aptitude Tests, the Cooperative English Test (C_2) and an abbreviated form of the Wechsler Adult Intelligence Scale were correlated with grade point averages. The corresponding coefficients of correlation were higher for boys than for girls in most instances. The best minimal predictor of grade point averages for both sexes was the comprehension section of the reading test. Grade point averages generally favored girls.

Intelligence and reading. Hoyt and Blackmore (31) studied the relationship of mental age based on the California Short Form Test of mental maturity to reading achievement measured in Grades I and II with the Ingraham-Clark Diagnostic Test and through Grade VII with the California Achievement Tests in Reading and the Iowa Tests of Basic Skills. They discovered that the 50 pupils who had been in the school throughout this period yielded comparable reading and mental age scores during the first three grades. However, at the fourth or fifth grades, reading achievement began to lag behind mental age and continued to do so throughout Grade VII. The investigators attributed this lag to lack of consistent reading instruction in the middle grades. A similar comparison was reported by Minkler (46) using the Stanford-Binet Intelligence Test and the Stanford Achievement Tests with 16 pupils having I.Q.'s above 100, enrolled in the school at least four consecutive years. She found that among bright children, reading age approaches and surpasses life age at the end of Grade I, then approaches mental age by Grade IV.

Reading, listening, language, intelligence, and scholastic achievement. Vineyard and Bailey (95) studied 114 second-semester college students using the Cooperative English Test for measuring reading achievement, the Listening Test of the STEP, and total scores from the ACE psychological examination to measure intelligence. Grade point averages for one year were called scholastic achievement. Based on intercorrelation and partial correlation coefficients, they concluded: 1) that reading ability, listening skill, and intelligence are highly interrelated and substantially related to scholastic achievement; 2) although reduced, each factor is related to the others when single ones are partialled out; and 3) with intelligence and reading held constant, listening is correlated with achievement; with reading and listening constant, intelligence is related to achievement; but with intelligence and listening controlled, reading is unrelated to achievement.

Differences in attainment in English (including reading comprehension) and arithmetic for pupils ten years of age from 16 primary schools in Wales were studied with respect to both intelligence and linguistic background by Lewis (38). School achievement tended to increase with intelligence but the differences in intelligence did not fully account for differences in achievement. In general, the attainment of pupils in bilingual schools was considerably lower in English but only slightly lower in arithmetic than was expected.

The Plattors and Sherwoods (59) investigated the relative superiority of the Pintner Verbal and the Pintner Non-Language tests using 12 classes at Grade VII. The mean I. Q.'s for the good readers were verbal, 113.6 and non-language 105.1 in contrast to those of poor readers with verbal of 74.5 and non-language of 84.5. The results support the conclusion that low I. Q.'s obtained by retarded readers may reflect the reading retardation rather than general inability to learn. Consequently the Pintner Verbal test alone is not a valid measure of learning capacity among these students. A nonverbal test is strongly recommended as an addition to the usual verbal intelligence test.

Sutton (82) collected reading test data, teachers' observations, and results from a personal data questionnaire about activities and interests for 23 of 210 pupils entering Grade I and continuing through Grade VI. All scored between the 90th and 100th percentile on the Metropolitan Readiness Tests. Two sample case studies are given showing excellent reading achievement records but differences in personal responses. From the group of case studies the author reached ten conclusions. Significant ones are: that precosity in reading may be detected early; that a school staff should establish "normal" reading achievement for each pupil, and that freedom of choice among many books with a stimulating home and school environment are essential for rapid progress.

Eye-movement norms. An extensive study of five characteristics of eye-movements is reported by Taylor, Frankenpohl and Pettee (84). The data were secured from 39 centers with many examiners and using more than 12,000 students in different geographical areas. The Reading Eye and recently prepared paragraphs ranging from 65-120 words for Grades I through college were used. Averages are presented for fixations and regressions per 100 words, span of recognition, duration of fixation and rate of reading. Separate averages revealed that girls were generally superior to boys, especially in the elementary grades. In the first three grades, averages reveal that the number of fixations, regressions, and the duration of fixation contribute to increase in reading rate. From Grades IV through VIII, duration of fixation was constant while reduction in number of fixations and regressions accompanied rate growth. Thereafter, the reduction in duration of fixation appeared to accompany increased rate.

Visual perception. Robinson, Mozzi, Wittick and Rosenbloom (68) followed the reading progress of 75 pupils through the first three grades and related it to scores on the Children's Perceptual Achievement Forms. Other factors such as intelligence, handwriting, and the need for visual care were assessed. Using a cutting score of 60 on the Perceptual Forms test and comparing the lower third with the upper two-thirds in reading achievement, the investigators found some of the relationships in Grades I and II to be statistically significant using the Chi-square statistic. At Grade III, however, there were no relationships between the scores on the Perceptual Forms and achievement in reading, handwriting, or the need for visual care. The Perceptual Forms were very slightly better than chance in identifying retarded readers. (Report at Grade I in 1958-59 summary.)

Vocabulary and reading progress. The hypothesis that an extensive vocabulary predisposes early reading success was tested by Potts (61). Using 197 children ages 72-78 months in England, he constructed a pictorial vocabulary test, used a word-reading test, a silent reading test, a phonic ability test, an intelligence test, and obtained socio-cultural ratings. Chronological age and net school attendance were considered also. The mean scores were tested for significance of differences, coefficients of intercorrelation were calculated, and the resulting matrix was subjected to factor analysis. The investigator found a "g" factor (which accounted for over half of the variance), a "maturity-schooling" factor and one including the three tests of reading. The conclusion was drawn that reading attainment and extent of vocabulary are different manifestations of general mental ability rather than distinct mutually contributive abilities.

Factors affecting interpretation. When students are asked to interpret materials, they bring to the selection their own set of values and perceptions. Thayer and Pronko (61) used as subjects 112 college sophomores to determine the effect of implied ethical and moral values on the reader's conceptualization of a fictitious character and his environment. Prior to reading the selection and answering 20 questions about it, a test of attitudes toward controversial moral issues and a measure of desirable and undesirable personality traits was administered. The passage of 200-250 words was fiction of an ambiguous type. The authors concluded that without description or with ambiguous description, readers ascribe desirable traits to a liked character and undesirable ones to a disliked character. But adding complex or somewhat undesirable moral issues "colors the readers' imagery and conceptualization," and that "functional cues may take precedence over structural cues in an ambiguous reading situation." Finally, they concluded that the reader appears to find it increasingly difficult "to extend and complete a story as moral and ethical implications become more evident."

Reading improvement and attitudes at the college level. Many effects of reading improvement have been noted by teachers, but Englander (21) constructed an instrument for securing more precise and comparable evidence. An open-ended questionnaire forced a choice of five answers to 16 questions. Each was rated from 1-5 relative to desirability. The questionnaire was administered to 46 subjects on the first and last day of a reading improvement course and to 30 controls enrolled in a public speaking class. Attitude toward reading as well as self-concept in relation to reading ability and as a student were significantly improved as compared with the control group.

Personality patterns and oral reading. Based on a review of the literature, Natchez (50) noted frustration as a characteristic of retarded readers when they read orally. She chose dependence, aggression, and withdrawal as symptoms of frustration. Subjects were 89 pupils in Grades V and VI—30 non-retarded and 59 retarded in reading. Using check-sheets for observation, cumulative records, opinions of the subjects' classmates, judgments of current teachers, and an experimental situation involving no reading the investigator compared the findings of the retarded with the non-retarded subjects using the T-test, and relied on coefficients of correlation for relationships. She concluded that "children who have difficulty with reading often experience considerable distress during the reading situation. This distress may evoke frustration reactions such as excessive dependence, aggression, and withdrawal." These reactions are greater than among pupils not retarded in reading. Furthermore, the manifestation of frustration in oral reading and in the experimental situation involving no reading were of a similar type.

Pre-school children's preferences in illustrations. Amsden (4) exposed 20 children at each of the ages three, four, and five to two sets of ten illustrations varying in amount and value of color or style of drawing. She found that these pupils significantly preferred: light tints and dark shades rather than brighter saturated colors, a photograph rather than black and white line drawing, those with several colors to those with fewer, and pictures placed on the right were chosen more often than those on the left.

Reading interests. One of the most extensive studies of children's interests in Grades III-VI yet published is Norvell's (53) report on 960,000 expressions of opinion of 24,000 pupils in New York state. The average number of reports per selection was 575 in Grade III and 615 in Grades IV-VI. Using a single score based on the percent marking a selection "very interesting," "fairly interesting," and "uninteresting," he ranked and listed 450 poems and 298 selections of various types of prose. He found similarities and differences in types of materials preferred by boys and girls. For example, boys prefer prose and girls interests in poetry rank higher. Individual selections show a wide spread in scores assigned to boys and girls. The degree of maturity reached, as represented by age, exerts a major influence on interests, with more rapid changes at younger ages. Rate of change in interests varies with materials. Interests in newspapers and periodicals are included. Expert opinion, represented by writers on children's literature, anthologists, librarians and teachers are challenged by the author's findings of the titles and authors liked best by children.

Hygiene of reading. Statistical and mathematical tables were studied for legibility by Tinker (88). Nine different but related studies used four pages of powers and roots of five-column tables. In each group there were 24-30 subjects. The conclusion was reached that grouping numerals by fives or tens in columns tends to produce more rapid location of powers and roots. The former is more effective for 6- and 8-point type. Little difference was found between one pica space and one pica with a rule between columns.

In the study of adult preferences in typography and design of book title-pages, Wrolstad (100) found that the 432 adults showed no significant preference for "right" versus "wrong" typography as they were classified by typographic designers. Young women showed greatest agreement and especially in the middle and high education groups they chose correct symmetric designs. While this study is classified as exploratory, it offers promise in improving communication through design.

The Teaching of Reading

A disproportionate number of investigations fall into this category for the current year.

Achievement and progress in reading. Pidgeon (58) reported on the achievement of 10,000 children attending different types of schools in England and Wales. Three samples were examined: ages 7-8, 10.5-11.5, and 14-15 years. Using the same test, he found that a small proportion of pupils in the middle and upper age groups scored at or below the mean score of the younger group. Sex differences in reading at the two upper age levels were not significant. Urban children surpassed rural children and the mean of those in county schools was higher than of those in voluntary schools. At age 14+, mean reading achievement of grammar school pupils exceeded that of modern schools, and the latter surpassed that of all-age schools. Likewise the mean reading scores of pupils attending single-sex schools exceeded that of mixed schools.

Worth (98) matched 33 pairs of third-grade pupils, half who were and half who were not promoted to Grade IV, with respect to sex, I.Q., and chronological age. He also designed a sociometric test and secured teachers' ratings on personality traits. Statistical comparison of the difference led him to conclude that the low achievers made approximately equal progress in achievement whether they were promoted or repeated third grade. Furthermore, non-promotion showed no adverse effect on social-personal development. However, he found a difference in gains made by the two groups on the California and Gates tests. Worth and Shores (90) analyzed the contents of the tests, compared the California with the Iowa reading tests, and pointed out possible reasons for the differences.

Growth in reading during a summer session was measured by Aasen (2), using 64 experimental and 32 control subjects in Grade IV. The experimental group was given a special program to stimulate reading books during the summer while the control group had none. The stimulated group read considerably more and made greater gains on the reading test than did the control group.

Comparative achievement. Scholl (74) used samples of English pupils at ages 7, 9, 11 and 13 years who took the reading and spelling sections of the Stanford Achievement Tests. The Burton children scored slightly above the test norms in spelling and word meaning, but below the norms in paragraph meaning. However, the differences were trivial despite the fact that English children entered school a year earlier than those on whom the norms were established.

Reading Readiness. Ploghoft (60) compared September scores in reading readiness of a group of children who had used reading readiness workbooks for nine weeks the previous spring with a group given an informal program. Although there were slight differences in mental ability favoring the experimental group, the mean scores of the two groups on the reading readiness test were not significantly different.

Durkin (18) prepared case studies of six pupils throughout first grade. She found that "the tendency to be passive and shy and to be lacking in self-confidence" seemed to deter reading progress. Some pupils appeared to make less progress due to lack of stimulation. There was considerable variation in the kind of teaching done with the same manual. The variations in I.Q. on three tests were marked.

Phonics. Bloomer (8) compared the progress of a first-grade class using a modified form of the Bloomfield phonic method with one using the basal reading program. The groups obtained similar scores on the Gates Reading Readiness Test, but after 24 weeks there was a significant difference in their achievement. The experimental group surpassed the control group in word recognition and in sentence reading, but there was no significant difference in paragraph meaning. He concluded that readiness for phonics is a function of the method of teaching.

Daniels and Diack (14) compared the progress of pupils using the "phonic word method" after one year in one junior school with those using three different adaptations of "mixed methods." They found statistically significant differences on tests of word recognition favoring the phonic method. The second experiment was in five infant schools, one of which used the phonic word method while each of the other four used variations of mixed methods. Again the phonic method proved to be superior in the area of word recognition but no other aspects were considered.

Of special interest is a study by Aaron (1) in which familiarity with phonic generalizations appropriate to Grades II and III were checked. Of the 293 subjects, 97 were prospective teachers or already teaching in the primary grades, 120 in the middle grades, and 76 were preparing for or teaching junior or senior high school. The mean score of correct items was about 57 percent of the total with a wide range. Furthermore, some of the phonics generalizations were more consistently understood than others. In addition, experienced teachers scored significantly higher than inexperienced teachers. The primary teachers, as a group, were no more familiar with these principles than were the teachers at the other levels.

Methods of teaching reading. Equalizing instructional time, Sheldon (78) compared the progress made by pupils at Grades IV and VI who had 90 minutes per week of instruction by TV with equated groups in which the same materials were used without TV. Although neither the names of the tests used nor the statistical data to support the conclusions were given in this report, he stated that at both grade levels, pupils taught by TV surpassed the control groups on the final tests of reading achievement. The differences were statistically significant at the .05 level at least. A questionnaire revealed an increase in positive attitudes toward reading among the sixth grade group taught by TV.

Durrell and four of his doctoral students (19) developed a team learning approach for adapting instruction to ability levels, learning rates and the needs in special skills in the middle grades of Dedham, Mass. Differentiation was attempted in reading, spelling, arithmetic and content subjects. "Reading achievement improved significantly only in grade six; there was no significant change... in grades four and five." The changes were attributed to a number of forces brought to bear simultaneously to improve learning.

Also at the middle grade level, Jenkinson and Lampard (32) outlined a model of the steps in problem solving, permitting nine teachers to vary the application as they wished. Based on the pre- and post-tests in reading and social studies, the STEP scores tended to show significantly improved performance when tests "demand critical thinking and reading." Although this was a pilot study for only five months, the results offer some promise as a technique of improving one aspect of reading.

Schiavone (73) described in detail a reading program at Grade VII, evaluated by the students' responses to a questionnaire. Each of the different types of instructional activities was enjoyed and in general they wished to continue to spend as much or more time on each. The order of preference in enjoyment was: library work, free reading, listening to stories read by the teacher, viewing films, reading newspapers and magazines, use of workbooks, and working in groups. However, the order of usefulness revealed that the first three choices were workbooks, the library and free reading. At the same level, Munro (48) experimented for four days with twelve classes reading content subject materials set up to present facts, relationships and generalizations. Pre- and post-tests were given and a retention test three weeks later. He concluded that in most of the differences analyzed, the three methods were equally effective, but wherever significant differences appeared, they favored the more meaningful methods. There was some evidence of transfer of the meaningful method from social studies to mathematics and no differences in retention.

Students achieving sub-college reading level received minimal instruction time but were stimulated to read more rapidly and widely, according to Weeks (96). In an academic year they improved markedly on informal and standardized tests of rate, but showed no consistent gains in vocabulary or comprehension. The control group, 26 students from freshmen composition classes, took the same tests and without stimulation, made small and inconsistent gains in all aspects of the tests.

Individualized reading and ability grouping. Sartain (70) chose ten classes of second grade pupils, randomly. Five classes began with the individualized program, using self-selection for three months, while the other five, grouped according to ability, used basal and supplementary books. Pre- and post-tests were administered, after which each group of five teachers reversed methods. After 56 school days, tests were administered to 234 pupils who completed the entire period. Teachers also appraised the strengths and weaknesses of the new approach. Analysis of variance revealed that both groups made greater gains during the first, as compared with the second, three-month period. Comparisons of achievement in individualized and basal programs revealed no significant differences except that those with lower I. Q.'s made greater gains in word recognition. He recommended adding individual conferences to the basal reader plan.

In Grades III through VI, Safford (69) identified seven classes, 183 children, taught by the individualized plan and compared their reading progress with national norms of the test for one year. He found that the majority of the pupils made less than average gains in reading during the experimental year. No significant differences were found between the gains of "superior" or "average" groups. Furthermore, individualized reading yielded about the same amount of improvement in vocabulary and comprehension.

Limited-range ability grouping of 156 pupils from Grades V through VIII was studied by Savard (72). Classes were divided into "upper range" and "lower range" groups and progress in reading and arithmetic for the experimental year was compared with that of the previous year when the same subjects were in heterogeneous groups. Based on statistical equivalence of Stanford Achievement Test scores, he found no advantage to upper range pupils in reading, but a dependable advantage in paragraph meaning for lower range classes who were grouped by ability.

Reading in the high school. Thornton (86) described a reading program for 40 college-bound seniors using pacers and workbooks as the major part of instruction. Pre- and post-tests given to these students and to a control group revealed significantly greater gains in rate in the experimental group while the control group had slightly higer means in vocabulary and comprehension. As a result of this evaluation, the program was changed to emphasize the last two aspects of reading and study skills for two summer classes. He found highly significant gains in all three aspects of reading and in attitudes of students taught to use higher skills.

About 20 percent of students who had six weeks of reading instruction in tenth grade were tested a year later to determine changes in rate of progress. Melton (44) reported numbers and percent of changes without statistical tests of significance. He analyzed the changes according to the quartile rank on an intelligence test. He found that 93 percent gained during the six weeks while 42 percent attained relatively higher percentiles one year later. No discernable differences appeared among the four quartile groups. Penty (56) reported again that "more than three times as many poor readers as good readers dropped out of school before graduation."

Improving college and adult reading. In his review of some research at the college level, Deverell (15) draws heavily upon reports from Canada and the United States. Among the nine conclusions, he points out that instruction designed to increase specific skills appears to do so without resulting in improvement in other areas. Hence his recommendation that freshmen programs include instruction in vocabulary, organizing reading and study procedures, critical reading and probably visual perception. He concludes, also, that gains obtained from special courses tend to improve academic achievement.

Evidence from four studies reveals that the most effective instructional techniques have not yet been established. Mayhew and Weaver (43) set up four conditions (combinations of materials and/or instruments) to test their effects on reading progress of college greshmen. They concluded that it is not profitable to make wide variations in levels of difficulty of materials in a single class and that neither the tachistoscope nor reading films are essential to develop skills, although they appear to add motivation. Kenworthy (37) reported on the instruction of adults in industry showing that emphasis on speed of reading resulted in improvement and maintenance (after one year) of rapid reading but without improvement or with slight loss in comprehension. The level of a difficulty of the SRA Book 3 was such that on some articles those who just read the titles answered the questions as well as those who read the selections. He reported that as the program progressed there was a shift away from "too much emphasis on increase in speed."

Miller (45) used a questionnaire to determine the types of instructional aids most frequently used in 233 college programs. He found that programs tended to be oriented around either workbooks or mechanical aids, although the latter seemed to be less satisfactory in some instances. He listed, in order of frequency of use for all students, the ten most widely selected workbooks, 14 types of supplementary materials, 14 mechanical aids, and seven standardized tests.

A different emphasis, group psychotherapy, was stressed by McDonald, Zolik, and Byrne (42). Although this was a pilot study, ten group-therapy sessions were given to one group and an extra hour of individualized reading per week to the other. The group subjected to therapy made better progress in reading, became more flexible in their approach, and generally improved on assessments of personality when compared with the control group. Likewise therapy seemed to contribute to better grades at the end of the semester and one year later.

Hill (29) chose to determine the contribution of vocabulary, phonetic analysis, understanding details, idea relationships and inferences to comprehension deficiency. His subjects represented "adequate readers" and "deficient readers." He used simple and multiple biserial correlation coefficients and beta regression weights. He concluded that recognition vocabulary made the largest contribution in discrimination to the five skill multiple, but that inferring tone and intent, and identifying idea relationships contributed somewhat to the differences in the two groups. Furthermore he discovered that limited intelligence, lower socio-economic home environment and certain personality variables also contributed to the "basic deficiency" in reading. Hence comprehension deficiency appears to be complicated by many factors. With different emphasis, Newton (51) found that the average reader differed substantially from the below average in the location of the secondary school attended, occupational status of parents, chronological age, and academic success in college. Only moderate significance was attached to sex ratio and size of secondary school.

A unique technique was devised by Raygor, Vance and Adcock (64) who asked each of 67 students referred to the Educational Skills Clinic to make 20 statements as descriptive as possible of his reading. The 781 statements were reduced to 285 and printed in a "True-False" questionnaire called the Diagnostic Reading Inventory (DRI). Judges separated items related to reading from those of feeling or attitudes toward reading. Further classification separated positive from negative feelings. An inverse cluster analysis identified students with similar descriptions. Predictive value of the DRI seems to offer some promise. Further analyses of data are anticipated.

Remedial reading. A wide range of factors have been associated with reading disability. Kawi and Pasamanick (36) examined records of pre-natal and birth histories of 205 male reading clinic cases and an equal number of matched controls to determine whether minimal cerebral injury was related to reading disability. They found some differences in socio-economic status which may have been related to the selection of cases. Of special interest was the significantly higher proportion of premature births among the retarded readers as compared with the controls. In addition, abnormalities of the prenatal period occurred with significantly greater frequency among retarded readers in comparison with controls. Of interest is the fact that visual and auditory difficulties did not appear to be associated with either reading disorders or the other factors under study.

Vernon's (94) critical analysis of studies of reading disability are particularly pertinent because in each instance she suggests steps to produce more dependable evidence. She concludes that the causes of reading disability may not be found until more is known about the fundamental nature of reading disability.

In the diagnosis of reading disability, the Wechsler Intelligence Scale for Children is frequently used. Two studies have been concerned with securing a special profile on this test to distinguish retarded readers from average readers. Sheldon and Garton (77) used deviation scores rather than scaled scores for 11 poor readers with an equal number of controls matched for age and sex. They found that the lowest subtest was "coding" which appears to have some similarity in process to word recognition, and the highest was Object Assembly. Dockerell (16) based his findings on 34 pupils and used the test norms for control. He found subtest of Coding, Arithmetic, Information were low while Picture Arrangement and Picture Completion scores were high. Comparisons are made with several previous studies of profiles on the same test.

Several reports deal with remedial instruction of poor readers. Hillman and Snowdon (30) used part-time untrained teachers working with small groups of pupils for two hours or more, depending primarily on a phonic approach. After 130 children had been taught, comparisons were made with matched groups without special help. Both average gains and comparisons with controls showed that significant reduction in retardation was achieved by this plan. Dunham (17) found that 20 retarded readers made significantly greater progress when given remedial education than did controls but the progress was not accompanied by corresponding improvement in attitudes toward reading. Roberts (65) combined three types of motivation and methods with 42 poor readers and concluded that his adaptation of the Fernald method was superior to the adaptation of the intrinsic method used by Gates and Schonell, and to the phonetic method similar to that reported by Monroe. The means of motivation differed also but seemed to be considered only incidental to the method of teaching, or as a means of explaining progress.

Pupils at Grade VIII who scored high on standardized reading tests, received excellent marks on report cards and had I. Q.'s at 110 or above, often have difficulty in reading according to Moe and Nania (47). The major handicaps were found to be in reading a chapter, in locational skills, and in flexibility of rate adapted to the level of difficulty of materials. Twelve pupils were given special instruction in their respective weak areas and marked improvement resulted.

Tamkin (83) reported that the reading achievement of 34 pupils admitted to a residential hospital for emotional disturbances was generally satisfactory. Arithmetic scores tended to be lower than reading.

Based on a summary of 18 selected references, Karlin (35) demonstrated the complexity of reading disability and suggested a few prerequisites for teachers, concluding that state certification of teachers was essential.

Predicting achievement from reading tests. Using coefficients of intercorrelation and multiple regression, Barrett (5) attempted to predict achievement of boys and girls in Grade IX from scores secured in Grade VIII on standardized tests of intelligence and locally constructed tests of arithmetic and spelling. Although a combination of the tests was a better predictor than any single test, the highest multiple coefficient was only .57. The Gates Reading Survey was judged least effective for both sexes and all courses, even though the local tests of arithmetic and spelling proved to be inadequate at the upper achievement levels.

Two studies deal with the use of reading tests for predicting grades of college freshmen. Robertson and Harrison (66) used the survey section of the Diagnostic Reading Test related to first-semester grade point averages of 152 students in different programs. Only vocabulary (which is related to intelligence) and total comprehension sections showed substantial correlation with grades. Analysis of expectancy reveals that the results of this test may be used as a rough but not a precise screening device in relation to general achievement. Aloia and Salinda (3) compared the sub-tests of the Cooperative English Test for 333 students with grades in English. They found the highest coefficient among the subtests to be in mechanics of expression, but all were moderate. It is clear that these tests of reading must be used with care as predictors of achievement at present.

At Grade XI, Ramsey (62) reported initial and final scores for 138 students on the Cooperative English Test of reading, the Word Attack section of the Diagnostic Test, and the California Test of Mental Ability. Statistically significant gains were found for all tests with brighter students making greatest gains. Coefficients of correlation between gains on tests of word attack and vocabulary, comprehension, and speed were insignificant, which he interpreted to mean that for these students, increasing ability in word attack does not result in general reading improvement. He suggested that other factors such as motivation and interest may have masked the expected relationships.

Tests. When new tests are made available, it is common to relate the scores to those of others used previously. Traxler (91) found that the split-half reliability of Form A of the Cooperative School and College Ability Test was quite satisfactory. Although there were substantial intercorrelations among the verbal, quantitative, and total scores, they were no higher than for other tests of this type. Both the verbal and total scores correlated at least as highly with reading comprehension scores as the reading tests correlate with each other. North (52) studied the reliabilities and difficulties of the Gates Primary and Advanced Primary Reading Tests at Grades II and III, for independent schools. He found that the tests were too easy for these able pupils, that the reliability coefficients ranged from .91 to .98 at second grade but were lower for the third grade. The relatively low sub-test scores (.80 and .79) suggest interpreting them with caution, while the average (.89) was more satisfactory.

The Gates Reading Diagnostic Tests and the Silent Reading Diagnostic Tests were administered to 20 pupils in Grades III through V by Murray and Karlsen (49). Paired scores on the sub-tests which purport to measure the same skills were compared. The assumption of concurrent validity between the tests was rejected in view of the highly significant differences in matched scores. The writers were unable to determine which test was more valid.

The Dominion Group Test of Reading Readiness —Long Form was reduced to shorter forms using the "subtest type" and the "Omnibus Type." Savage (71) related scores on these short forms secured in October to the scores on paragraph reading obtained in May using 1201 pupils in 47 schools of Ontario. He concluded that the validity coefficients were low and that the short-form tests were invalid for this population in which other contributing factors, such as intelligence, were not controlled.

Pauk (55) was concerned with selected characteristics of boys compared with girls who scored above and below average on the Cooperative Reading Test at ninth grade. He found that the language-type tests contributed about 80 percent to the total battery in differentiating good from poor girl readers but only about 37 percent for boys. The reverse was true for the Abstract Reasoning Test and the Numerical Ability Test; for girls, about 11 percent in contrast to 43 percent for boys.

Curr and Gourlay (13) examined the effects of practice on tests of "mechanical reading," reading comprehension and arithmetic with pupils 9 and 10 years of age. Assuming a normal mean gain over periods of one, three, and six months, retests showed 10.1, 18.2 and 26.9 months respectively for average pupils. Practive effect was greatest in comprehension, less in "mechanical reading," and slight in arithmetic. Implications for interpreting frequent retests of retarded readers were pointed out.

At the college level, Halfter and Douglass (27) developed a Commerce Reading Comprehension Test. The details of selecting items, refining them, developing tentative norms, testing for validity and reliability, are reported. Furthermore, coefficients of correlation with standardized reading tests as well as certain special tests for accountants reveal high relationships.

BIBLIOGRAPHY

1. Aaron, I. E. "What Teachers and Prospective Teachers Know about Phonics Generalizations," Journal of Educational Research, LIII (May, 1960), 323-30.

   Reports results of a study among 293 teachers and prospective teachers to assess their knowledge of second and third grade phonic generalizations and the relationship between extent of knowledge and amount of teaching experience, educational assignment of experienced teachers, and preferences of prospective teachers.

2. Aasen, Helen B. "A Summer's Growth in Reading," Elementary School Journal, LX (November, 1959), 70-74.

   Compares the growth in reading during the summer vacation months as measured by the Stanford Reading Test, Form J, of an experimental and a control group of fourth-grade children, the former having been stimulated to reading during the summer, whereas the latter was not

3. Aloia, Alex D. and Salinda, Juan F. "A Correlation Study Between Grades in English and Cooperative English Test Scores of College Freshmen," California Journal of Educational Research, XI (January, 1960), 7-13, 47.

   Reports the results of a study of the correlation, if any, "between English 11 grades and scores on the Cooperative English Test (Higher Level, Form Z)" to determine the predictive value of the latter.

4. Amsden, Ruth Helen. "Children's Preferences in Picture Story Book Variables," Journal of Educational Research, LIII (April, 1960), 309-12.

   Reports findings of a study involving 60 boys and girls, aged three to five, from three different nursery schools, to discover their

"preferences in illustrations and story subjects."

5. Barrett, Harry O. "The Predictive Efficiency of Grade 8 Objective Tests in Terms of Grade 9 Achievement," Ontario Journal of Educational Research, II (April, 1960), 101-07.

   Presents and interprets the correlations of scores on reading and arithmetic tests with marks on Christmas examinations in four academic fields by 2,510 ninth-grade pupils to determine the predictive value of the objective tests used.

6. Bliesmer, Emery P. "Review of Recent Research on College and Adult Reading," Starting and Improving College Reading Programs. Eighth Yearbook of the National Reading Conference. Fort Worth, Texas: The Texas Christian University Press, 1959, 171-92.

   Summarizes the results of 82 research reports, under the following headings: "reading programs," "reading and study, skills and habits--status indications," "influence of skills and habits," "factors influencing skills and habits," "tests," "eye-movements and controlled reading."

7. Bloom, Benjamin S. "Appraising Reading Progress Under Different Patterns of Grouping," Reading Instruction in Various Patterns of Grouping. Supplementary Educational Monographs. No. 89. Chicago: The University of Chicago Press. 1959. 180-84.

   Distinguishes between exploratory and demonstration research and discusses requisites of the latter as applied to the study of different patterns of grouping.

8. Bloomer, Richard H. "An Investigation of an Experimental First Grade Phonics Program," Journal of Educational Research, LIII (January, 1960), 188-93.

   Compares the progress of two groups of first-grade pupils whose average scores were about the same on the Gates Reading Readiness Test, one group acquiring a sight vocabulary before phonics was introduced and the other receiving training in phonics before beginning to read.

9. Blum, Eleanor. "Paperback Book Publishing: A Survey of Content," Journalism Quarterly, XXXVI (Fall, 1959), 447-54.

   Reports the results of a detailed study of the content of 4,868 original works and reprinted titles which were in print in 1955.

10. Bush, Chilton R. "A System of Categories for General News Content," Journalism Quarterly, XXXVII (Spring, 1960), 206-10.

    Presents a detailed classification scheme of categories for general newspaper content so defined as to be mutually exclusive and reports the results of efforts to validate them.

11. Clark, Willis W. "Boys and Girls-Are There Significant Ability and Achievement Differences?" Phi Delta Kappan, XLI (November, 1958), 73-76.

    Compares the mental ability (eight measures) and achievement (six measures including reading) of boys and girls (grades three, five and eight) drawn at random from a nation-wide stratified sample for the country as a whole.

12. Cornwell, Elmer E., Jr. "Presidential News: The Expanding Public Image," Journalism Quarterly, XXXVI (Summer, 1959), 275-83.

    Presents an analysis of the front page news in a large daily and a medium-sized daily over a period of 70 years to determine trends in the extent of news published relating to the President and his activities.

13. Curr, W. and Gourlay, N. "The Effect of Practice on Performance in Scholastic Tests," The British Journal of Educational Psychology, XXX (June, 1960), 155-67.

    Reports the extent of "practice effect," as measured by tests "in mechanical reading, reading comprehension and arithmetic," given "at varying intervals of time to groups of 9-10-year-old children in four primary schools." (The second paper in a symposium on "The Diagnosis and Remedial Treatment of Reading Difficulties.")

14. Daniels, J. C. and Diack, Hunter. "The Phonic Word Method," The Reading Teacher, XIII (October, 1959), 14-21.

    Includes a brief report of two experiments in England which aimed to determine the comparative efficiency of the "phonic word method" and six other approaches to reading.

15. Deverell, A. F. "Are Reading Improvement Courses at the University Level Justified?" Invitational Conference on Educational Research, The Canadian Educational Association, Saskatoon, September 15 and 16, 1959. 19-27.

    Presents a partial survey of research relating to Reading courses in colleges and universities in Canada and the United States.

16. Dockrell, W. B. "The Use of Wechsler Intelligence Scale for Children in the Diagnosis of Retarded Readers," The Alberta Journal of Educational Research, VI (June, 1960), 86-91.

    Reviews results of related studies and reports the results of a study involving 34 retarded readers, ages 8.2 to 14.9 years, for whom a full W.I.S.C. profile was available to determine the diagnostic significance of the profile as a whole and of its sub-parts.

17. Dunham, J. "IV-The Effects of Remedial Education on Young Children's Reading Ability and Attitude to Reading," The British Journal of Educational Psychology, XXX (June, 1960), 173-75.

    Reports the results of an experiment with twenty severely retarded children of about av-

erage intelligence in which remedial reading lessons were given once or twice weekly for six months to determine their effectiveness in terms of reading ability and attitude toward reading. (The fourth paper in a symposium on "The Diagnosis and Remedial Treatment of Reading Difficulties.")

18. Durkin, Dolores. "A Case-Study Approach Toward an Identification of Factors Associated with Success and Failure in Learning to Read," California Journal of Educational Research, XI (January, 1960), 26-33.

Reports findings in six case studies at the first-grade level, the ultimate aim being to secure "a better understanding of the number and kinds of factors that seem to be associated with both success and failure in learning to read."

19. Durrell, Donald D., Manning, John C., McHugh, Walter J., Rochfort, George B., Jr., Scribner, Harvey B. "Adapting Instruction to the Learning Needs of Children in the Intermediate Grades," Journal of Education, Boston, CXLII (December, 1959), Entire issue.

Summarizes the results of four cooperative doctoral studies which aimed to check the assumption that learning, including progress in reading, is most effective when it is adapted to the learning needs of children (803 pupils in grades 4, 5 and 6).

20. Eames, Thomas H. "Some Neural and Glandular Bases of Learning," Journal of Education, Boston, CXLII (April, 1960), Entire issue.

An excellent summary of pertinent medical literature bearing on the task of learning to read and reading disability.

21. Englander, Meryl E. "Changes in Affect Attributable to Instruction in Reading Improvement at the College Level," Journal of Educational Research, LIII (February, 1960), 231-36.

Presents the results of an exploratory study involving 46 students to determine through the use of an Instrument for Measuring Student's Attitudes, the effect of participation in a reading improvement program on their attitude toward reading and allied subjects.

22. Entwisle, Doris R. "Evaluations of Study-Skills Courses: A Review," Journal of Educational Research, LIII (March, 1960), 243-51.

Summarizes 22 evaluation reports of study-skill courses in respect to "kind of student," "criterion used," "baseline used," "methods used to correct for bias," and "performance in excess of baseline" both immediate and follow-up.

23. Fisher, Helen. "Teacher Differences in Professional Reading," Educational Administration and Supervision, XLIV (September, 1958), 282-89.

Summarizes the results of information secured from 50 elementary school teachers to determine their patterns of professional reading and the factors that influenced their reading.

24. Goldhor, Herbert. "Are the Best Books the Most Read?" The Library Quarterly, XXIX (October, 1959) 251-55.

Compares the circulation and value category of books in a sample of hard-cover books during three different periods since 1900.

25. Gray, William S. "Summary of Investigations Relating to Reading July 1, 1958 to June 30, 1959," Journal of Educational Research, LIII (February, 1960), 203-22.

Summarizes major findings of 120 investigations relating to reading; includes an annotated bibliography.

26. Hage, George S. "Anti-Intellectualism in Press Comment: 1828 and 1952," Journalism Quarterly, XXXVI (Fall, 1959), 439-46.

Presents the results of an analysis of newspaper content in the Eisenhower versus Stevenson and the Jackson versus Adams campaigns, showing the extent of the anti-intellectualism content in each.

27. Halfter, Irma T. and Douglass, Frances M. "Measurement of College Level Reading Competence in a Content Area," Journal of Educational Research, LIII (February, 1960), 223-30.

Describes the steps taken in developing a test of college level reading competence in the field of business to determine if the skills required involve more than those usually regarded as comprehension in reading.

28. Haskins, Jack B. "Pre-testing Editorial Items and Ideas for Reader Interest," Journalism Quarterly, XXXVII (Spring, 1960), 224-30.

Describes a method of testing the reader interest value of editorials, items and ideas, presents data concerning the reliability and validity of "title rating," considers ways in which editors and researchers might use the technique.

29. Hill, Walter R. "Factors Associated with Comprehension Deficiency of College Readers," Journal of Developmental Reading, III (Winter, 1960), 84-93.

Analyzes the scores made by 164 university freshmen on a battery of tests and rating scales "to determine the extent to which selected measures of reading skills, verbal and non-verbal intelligence, personality, experiential background and academic status" would discriminate between freshmen with adequate and deficient reading comprehension.

30. Hillman, H. H. and Snowdon, R. L. "III—Part-Time Classes for Young Backward Readers," The British Journal of Educational Psychology, XXX (June, 1960), 168-72.

Reports the results of a study to determine

whether part-time teachers without special training could "significantly improve the reading achievement of their pupils" when working with small groups of backward children and using prescribed procedures. (The third paper in a symposium on "The Diagnosis and Remedial Treatment of Reading Difficulties.")

31. Hoyt, Jeanne S. and Blackmore, Dorothy S. "Fifty Seventh Graders: A Comparison of Their Reading Achievement and Expected Achievement in Grades One Through Seven," *Journal of Educational Research*, LIII (January, 1960), 163-71.

    Compares the achievement at each grade level of 50 seventh-grade pupils who had been in the school system since the first grade with their expected achievement as indicated by their intelligence scores determined in the seventh grade.

32. Jenkinson, Marion D. and Lampard, Dorothy M. "Problem Solving and Achievement," *The Alberta Journal of Educational Research*, V (September, 1959), 166-72.

    Compares the scores on reading and social studies tests given to three experimental and three control groups in each of three grades (IV, V and VI) before and after a period of training in problem solving to determine the influence of such training on improvement in these subjects.

33. Josephina, C. J. S., Sister. "Survey of the Research Related to the Reading Ability of the Gifted," *Journal of Educational Research*, LIII (February, 1960), 237-39.

    Summarizes "some of the major findings in reading performance as related to the gifted pupil, especially in the elementary grades.

34. Karlin, Robert. "Research in Reading," *Elementary English*, XXXVII (March, 1960), 177-83.

    Summarizes important conclusions from 46 investigations relating to "reading readiness," "phonics," "interests," "comics and television," "content fields," "Individual reading," "machines."

35. Karlin, Robert. "Who Are Treating Our Disabled Readers?" *The Reading Teacher*, XIII (April, 1960), 288-92.

    Reviews briefly what research says about the causes of reading failures to determine the extent "to which real competence in teaching is necessary if children are to be served."

36. Kawi, Ali A. and Pasamanick, Benjamin. *Prenatal and Paranatal Factors in the Development of Childhood Reading Disorders.* Monographs of the Society for Research in Child Development Serial No. 73, Vol. 24, No. 4, Lafayette, Indiana: Child Development Publications, Purdue University, 1959.

    Compares "the prenatal and paranatal records of 372 white male children with reading disorders born in Baltimore between 1935 and 1945," with the records of a similar number of matched controls to identify possible factors related to the development of reading disorders.

37. Kenworthy, C. W. "An Evaluation of the Results of Instruction and Practice in the Techniques of Better Reading," *Journal of Developmental Reading*, II (Summer, 1959), 11-16.

    Reports the methods and results of a reading improvement program, involving 12 ninety-minute sessions, given to 150 members of an industrial plant and providing measures of gains a year after the training discontinued.

38. Lewis, D. G. "Differences in Attainment Between Primary Schools in Mixed-Language Areas: Their Dependence on Intelligence and Linguistic Background," *The British Journal of Educational Psychology*, XXX (February, 1960), 63-70.

    Bases conclusions on the results of attainment tests in English (comprehension and usage) and arithmetic (mechanical problems) "given to all the ten-year-old pupils from 16 primary schools in mixed language areas in Wales"; "a non-verbal test of intelligence and a linguistic background questionnaire were also given."

39. Lewis, Ralph. "Books That Germans Are Reading About America," *The Library Quarterly*, XXIX (October, 1959), 246-50.

    Reports the results of an analysis of the books relating to America read most frequently by Germans to determine the possible effect that it might produce on the reader's attitude toward America.

40. Lockwood, J. Bryce. "Research on Problems in Reading Science," *School Science and Mathematics*, LIX (October, 1959), 551-56.

    Summarizes the findings of nineteen research studies relating to "the development of scientific vocabulary,"; "the level of reading difficulty of textbooks for science", and "the ability of science teachers to estimate the reading difficulty of science textbooks."

41. McCullough, Constance M. "Implications of Research on Children's Concepts," *The Reading Teacher*, XIII (December, 1959), 100-07.

    Summarizes the findings of 39 studies under "what a concept is," "how concepts are formed," "factors influencing concept development," "what teachers can do."

42. McDonald, Arthur S., Zolik, Edwin S., and Bryne, James A. "Reading Deficiencies and Personality Factors," *Starting and Improving College Reading Programs*, Eighth Yearbook *National Reading Conference*. Fort Worth, Texas: The Texas Christian University Press, 1959, 89-98.

Reports the methods and results of an uncontrolled pilot study including college students who were achieving below expectancy on the basis of high school grades and scores on entrance examinations to assess the possibility of contributions from psychotherapy to a college improvement reading program.

43. Mayhew, Jean B. and Weaver, Carl H. "Four Methods of Teaching Reading Improvement at the College Level," Journal of Developmental Reading, III (Winter, 1960), 75-83.

Compares the gains in reading skills by 96 beginning students "in regularly scheduled classes under four different methods of teaching."

44. Melton, Howard B. "Reading Retention of Eleventh-Grade Students," Journal of Developmental Reading, II (Summer, 1959), 26-33.

Studies the effectiveness of a reading improvement program given to 66 tenth-grade students by comparing their reading ranks at the beginning and ending of the program and approximately a year later.

45. Miller, Lyle L. "Current Use of Workbooks and Mechanical Aids," Starting and Improving College Reading Programs, Eighth Yearbook of the National Reading Conference. Fort Worth, Texas: The Texas Christian University Press, 1959, 67-75.

Summarizes responses from 233 colleges relating to "growth or decline in reading program," "college credit," "time spent on reading instruction," "utilization of workbooks," "materials to supplement workbooks," "mechanical aids."

46. Minkler, Rachel S. "Developmental Patterns in the Fundamental Skills of the Elementary School Programme," Ontario Journal of Educational Research, II (October, 1959), 45-50.

Compares with their mental and chronological ages, the reading and arithmetic ages of 16 children in a small private elementary school at each successive grade level from I to VI, as determined through the use of standardized tests.

47. Moe, Iver L. and Nania, Frank. "Reading Deficiencies Among Able Pupils," Journal of Developmental Reading, III (Autumn, 1959), 11-26.

Reports the results of the diagnosis and corrective training in reading of twelve pupils with excellent report-card marks, high scores on two different standardized reading tests, and intelligence quotients of 110 or better.

48. Munro, Barry C. "Meaning and Learning," Alberta Journal of Educational Research, V (December, 1959), 268-81.

Reports the results of a controlled experiment, including twelve seventh-grade classes, "to determine the learning outcomes" from the study of assigned materials, "arranged in an ascending order of meaningfulness"—facts, relationships, generalizations.

49. Murray, Carol-Faith and Karlson, Bjorn. "A Concurrent Validity Study of the Silent Reading Tests and the Gates Reading Diagnostic Tests," The Reading Teacher, XIII (April, 1960), 293-94, 296.

Compares the means and standard deviations of the scores made by twenty children with reading disabilities on two tests which purport to measure the same skills to determine if they yield equivalent distributions of scores.

50. Natchez, Gladys. Personality Patterns and Oral Reading. New York: New York University Press, 1959.

Reports a controlled experiment including 59 retarded readers and 30 control subjects "to determine the relationship between a child's reaction during a reading situation and his reactions in general."

51. Newton, Eunice Shaed. "Empirical Differences Between Adequate and Retarded Readers" The Reading Teacher, XIII (October, 1959), 40-44.

Compares 176 adequate readers and 118 retarded readers at the college level in respect to "location and size of secondary school, occupational status of parents, sex, chronological age, and earned semester averages."

52. North, Robert E. "Difficulty and Reliability of the Gates Primary and Advanced Primary Reading Tests, 1958 Edition, for Independent School Pupils," 1959 Fall Testing Program in Independent Schools and Supplementary Studies, Educational Records Bulletin No. 76. New York: Educational Records Bureau, 1960, 45-51.

Bases conclusions on test records secured from 369 second and 329 third-grade pupils in 17 schools.

53. Norvell, George W. What Boys and Girls Like to Read. Morristown, N. J.: Silver Burdett Company, 1958.

Summarizes and interprets "expressions of opinions" from more than 24,000 children in grades III to VI concerning 1576 literary selections.

54. Ong, Jin, Schneider, Kenneth, and Moray, Joseph. "Reading Ability and Perimetric Visual Field," California Journal of Educational Research, XI (March, 1960), 61-67.

Reports the results of a study involving 155 children, ages about 11, to determine the characteristics of their "perimetric peripheral visual fields" and to relate the sizes of fields to various factors such as sex, IQ's, power reading and speed of reading.

55. Pauk, Walter J. "Are Present Reading Tests Valid for both Boys and Girls?" Journal of

Educational Research, LIII (March, 1960) 279-80.

Compares the scores of 150 ninth-grade boys and girls on seven tests other than reading to determine specific skills and attitudes which differentiate in the case of each sex, above and below average readers, as measured by the Cooperative Reading Comprehension Test, C1 (Lowest Level).

56. Penty, Ruth C. "Reading Ability and High School Drop-Outs," National Association of Women Deans and Counselors Journal, XXIII (October, 1959), 11-15.

Presents a general summary of a study entitled Reading Ability and High School Drop-outs by Ruth Penty, New York: Bureau Publications, Teachers College, Columbia University, 1956 (included in the 1955-56 annual summary).

57. Perrodin, Alex F. "Televiewing, Reading Habits, and Children's Social Values," Elementary English, XXXVII (February, 1960), 86-90.

Reports data from 352 pupils in grades IV to VII inclusive "to determine possible relationships between certain television and reading habits of children and selected behavior traits"—cooperation, friendliness, leadership, responsibility.

58. Pidgeon, D. A. "A National Survey of the Ability and Attainment of Children at Three Age Levels," The British Journal of Educational Psychology, XXX (June, 1960), 124-33.

Summarizes the results of tests in reading, arithmetic and non-verbal ability given to a sample of some 10,000 children from "maintained or assisted schools" in England and Wales in three age groups (7, 10, and 14) to discover existing standards and thus provide a basis for future comparisons, for determining amount of backwardness in each age group and for comparing the attainment of pupils in different types of schools.

59. Plattor, Emma E., Plattor, Stanton D., Sherwood, Clarence and Sherwood, Sylvia. "Relationship between Reading Retardation and the Measurement of Intelligence," Personnel and Guidance Journal, XXXVIII (September, 1959), 49-51.

Reports scores made by "on grade" and retarded seventh graders on verbal and nonverbal intelligence tests to determine the validity of the assumption that "low intelligence quotients obtained by retarded readers may reflect their reading retardation rather than a basic inability to learn."

60. Ploghoft, Milton H. "Do Reading Readiness Workbooks Promote Readiness?" Elementary English, XXXVI (October, 1959), 424-26.

Summarizes the results of a partially controlled experiment, including a total of 55 kindergarten children, to determine the effectiveness of the use of readiness workbooks as measured by improvement during the training period and summer in scores made on the Metropolitan Readiness Test.

61. Potts, Eric. "A Factorial Study of the Relationship Between the Child's Vocabulary and His Reading Progress at the Infants' Stage," (Abstract of Thesis submitted for the degree of M. Ed. in the University of Manchester, April, 1959) The British Journal of Educational Psychology, XXX (February, 1960), 84-86.

Reports the results of a study including 107 boys and 90 girls between 72 and 78 months of age from eight infant schools to test the hypothesis that "an extensive vocabulary predisposes a child to early success in reading."

62. Ramsey, Wallace. "An Analysis of Variables Predictive of Reading Growth," Journal of Developmental Reading, III (Spring, 1960), 158-64.

The results of a statistical study based on test scores from 138 eleventh grade students "to determine which of several variables were predictive of success in improving various important reading abilities," particularly speed, comprehension and vocabulary.

63. Rankin, Earl F. J. "The Cloze Procedure—Its Validity and Utility," Starting and Improving College Reading Programs, Eighth Yearbook of the National Reading Conference. Fort Worth, Texas: The Texas Christian University Press, 1959, 131-44.

Summarizes "evidence concerning the empirical validity of the procedure as a technique for measuring readability, intelligence, pre-reading knowledge, comprehension."

64. Raygor, Alton V., Vance, Forrest L., and Adcock, Donna. "The Diagnosis and Treatment of College Reading Difficulties Using Patterns of Symptomatic Statements," Journal of Developmental Reading, III (Autumn, 1959), 3-10.

Reports the results of a search for a method of developing "symptomatic syndromes independent of preconceived notions as to what the basic kinds of reading disability might be" through the use of the student's own self descriptions.

65. Roberts, Geoffrey R. "V—A Study of Motivation in Remedial Reading," The British Journal of Educational Psychology, XXX (June, 1960), 176-79.

Reports the results of experiments with poor readers to determine the advantages and limitations of three methods of motivating effort. (The fifth paper in a symposium on "The Diagnosis and Remedial Treatment of Reading Difficulties".)

66. Robertson, Malcolm H. and Harrison, Mildred M. "Reading Skill As a Predictor of College Achievement," Journal of Educational Re-

search, LIII (March, 1960), 258-62.

Presents correlations between the scores made by 152 college freshmen on the Diagnostic Reading Test: Survey Section with semester grade point averages to test the hypothesis "that scores on a standardized reading test show a predictive relationship to college grades."

67. Robinson, H. Alan and Dramer, Dan S. "High School Reading - 1958," *Journal of Developmental Reading*, III (Winter, 1960), 94-105.

The first of a series of annual summaries of the professional literature on junior and senior high school reading, including pertinent scientific studies.

68. Robinson, Helen M., Mozzi, Lucille, Wittick, Mildred Letton, and Rosenbloom, Alfred A. "Children's Perceptual Achievement Forms: A Three Year Study," *American Journal of Optometry and Archives of American Academy of Optometry*, XXXVII (May, 1960), 223-37.

Reports the results of a study to determine the value of the Children's Perceptual Achievement Forms as a predictor of reading achievement, using scores on various tests given to those of 87 first grade pupils in 1956-57 who remained in school for the three years.

69. Safford, Alton L. "Evaluation of an Individualized Reading Program," *The Reading Teacher*, XIII (April, 1960), 266-70.

Analyzes the scores made on the reading section of the California Test Battery by pupils in seven classes representing Grades III to VI inclusive and who had been taught by the individualized teaching technique to determine extent of growth as compared with that of pupils taught by other techniques in the school system.

70. Sartain, Harry W. "The Roseville Experiment With Individualized Reading," *The Reading Teacher*, XIII (April, 1960), 277-81. (Also see related bibliography pp. 262-64.)

Reports a carefully planned study including ten second-grade classes to compare the progress of pupils taught by individualized self selection and by basic readers and organized in ability groups with a variety of supplementary books.

71. Savage, H. W. "Validity of the Dominion Group Test of Reading Readiness—Short Form, and Differences Among Groups of Pupils Tested," *Ontario Journal of Educational Research*, II (October, 1959), 63-70.

Summarizes and interprets scores on two types of short forms of the test given in October, 1955 and on the Dominion Achievement Test in Silent Reading given in 1956.

72. Savard, William G. "An Evaluation of an Ability Grouping Program," *California Journal of Educational Research*, XI (March, 1960), 56-60.

Compares the progress of more than 1200 pupils in grades IV to VIII under conditions of "limited range grouping" with that of the previous year under usual heterogeneous class groupings; also explores the relationship between intelligence and the effects of limited range grouping.

73. Schiavone, James. "A Seventh-Grade Reading Enrichment Program," *Journal of Developmental Reading*, III (Winter, 1960), 106-14.

Summarizes responses of students to a questionnaire requesting their evaluation of the activities pursued in a reading enrichment program and the time devoted to them.

74. Scholl, Geraldine T. "The Reading and Spelling Achievement of a Group of English Children," *University of Michigan School of Education Bulletin*, XXXI (April, 1960), 105-07.

Compares the achievement of 150 children in Burton, England with norms on the Stanford Achievement Test.

75. Schramm, Wilbur, Lyle, Jack, and Parker, Edwin B. "Patterns in Children's Reading of Newspapers," *Journalism Quarterly*, XXXVII (Winter, 1960), 35-40.

Summarizes information secured from 200 to 400 pupils per grade in grades I to VI and in the eighth, tenth and twelfth grades in representative schools in San Francisco, concerning various aspects of their reading, with special reference to newspaper reading, and their use of other mass media.

76. Schubert, Delwyn G. "Do Teachers Read About Reading?" *California Journal of Educational Research*, XI (March, 1960), 94-96.

Summarizes the responses of 132 experienced teachers in California and Washington to six questions relating to grade taught, amount of teaching experience, number of professional books on reading owned, value attached to such books, and number of magazines subscribed for that carry articles on reading.

77. Sheldon, M. Stephen and Garton, Jeanette. "A Note on a WISC Profile for Retarded Readers," *Alberta Journal of Educational Research*, V (December, 1959), 264-67.

Summarizes test scores "to check and supplement the findings of Altus (1956) on the differential performance of retarded readers on the subtests of the Wechsler Intelligence Scale for Children."

78. Sheldon, William D. "Television and Reading Instruction," *Education*, LXXX (May, 1960), 552-55.

Describes procedures and findings in a partially controlled experiment including "average readers" in grades four and six to determine the value of the three TV lessons per week as part of the total reading program.

79. Staiger, Ralph C. "Language Arts Research, 1958," *Elementary English*, XXXVI (Novem-

ber, 1959), 502-10.

Lists 284 research studies relating to reading reported as completed or in progress during 1958. A corresponding list for 1956 appeared in the April, 1957 issue of Elementary English.

80. Standish, E. J. "Readiness to Read," Educational Research, II (November, 1959), 29-38.

Discusses the findings and implications of twenty-one studies relating to various aspects of reading readiness: "the task of the teacher of beginners," "ascertaining stages of reading development," "the use of reading tests for diagnostic purposes," "the value of reading readiness tests," "teachers' attitudes towards and recognition of reading readiness."

81. Stinson, Pairlee J. and Morrison, Mildred M. "Sex Differences Among High School Seniors," Journal of Educational Research, LIII (November, 1959), 103-08.

Presents an analysis of test scores for 36 boys and 33 girls to determine "the relationships of the subtests of the Differential Aptitude Tests and the Cooperative English, C 2, to grade-point average and to an abbreviated form of the Wechsler Adult Intelligence Scale."

82. Sutton, Rachel S. "Variations in Reading Achievement of Selected Children," Elementary English, XXXVII (February, 1960) 97-101.

Describes the procedures and presents examples of findings in a study of variations in reading achievement of children over a seven-year period who scored high in reading readiness in the kindergarten; sought also to identify environmental influences that tend to foster individuality and independence in reading.

83. Tamkin, Arthur R. "A Survey of Educational Disability in Emotionally Disturbed Children," Journal of Educational Research, LIII (April, 1960), 313-15.

Summarizes the results of a study including 22 boys and 12 girls, mean ages 8. 7, and 9.4 respectively, who had been hospitalized for treatment of emotional disorders, to ascertain the frequency of educational disability in reading and arithmetic and relationships, if any.

84. Taylor, Stanford E., Frackenpohl, Helen and Pettee, James L. Grade Level Norms for the Components of the Fundamental Reading Skill, Educational Developmental Laboratories Research and Information Bulletin No. 3. New York: Educational Developmental Laboratories, Inc., 1960.

Norms for five components of reading for grades one to college inclusive, based on eye-movement records from 12,143 subjects in non-homogeneously grouped classes in schools widely distributed throughout the country.

85. Thayer, Leo O. and Pronko, N. H. "Factors Affecting Conceptual Perception in Reading," Journal of General Psychology, LXI (July, 1959), 51-59.

Describes the methods and results of a study among 112 college sophomores to determine how implied ethical and moral values alter a "reader's perceptions and conceptualizations of the people and ideas he was reading about."

86. Thornton, Cecil M. "Two High School Reading Improvement Programs," Journal of Developmental Reading, III (Winter, 1960), 115-22.

Describes efforts, including a partially controlled experiment, to improve high school reading, and presents a statistical analysis of scores before and after training to identify elements of strength and weakness in the procedures followed.

87. Tinker, Miles A. "Print for Children's Textbooks," Education, LXXX (September, 1959), 37-40.

Summarizes the results of ten investigations relating to the legibility of print under the following headings: "size of type," "length of line," "leading," "style of type face," "quality and color of paper," "color of paper and ink."

88. Tinker, Miles A. "Legibility of Mathematical Tables," Journal of Applied Psychology, XLIV (April, 1960), 83-87.

Reports findings of an experiment including 246 subjects to determine "the effect of type size, arrangement of numerals in columns, and space versus space plus a rule between columns on the speed of locating powers and roots in mathematical tables."

89. Townsend, Agatha. "Another Look at Reading Interests," The Reading Teacher, XIII (April, 1960), 297-98, 301-02.

Reviews trends in amount of research relating to reading interests, points out significant weaknesses in reported research, and suggests the nature of needed research in the field.

90. Traxler, Arthur E. "Recent Findings and Trends in Research on Reading," The Reading Teacher, XIII (December, 1959), 90-99.

Summarizes the major findings of research on reading, 1953-57, under such headings as readiness, interest, reading and other subjects, phonics, tests, speed of reading, visual defects, auditory defects, typography, diagnosis, remedial and corrective work.

91. Traxler, Arthur E. "Some Independent School Results on the Cooperative School and College Ability Tests, Levels 2-5," 1959 Fall Testing Program in Independent Schools and Supplementary Studies, Educational Records Bulletin, No. 76. New York: Educational Records Bureau, 1960, 57-64.

Presents correlations of scores "on the Schools Ability Test with comprehension on three reading tests for several small groups of independent school pupils."

92. Traxler, Arthur E. and Jungeblut, Ann, with the assistance of the Educational Records Bureau Staff. Research in Reading During Another Four Years: Summary and Bibliography. Educational Records Bulletin, No. 75. New York: Educational Records Bureau, May, 1960. Entire issue.

Summarizes the major findings of 438 scientific studies relating to reading published between July 1, 1953 and December 31, 1957, includes annotated bibliographical references.

93. Vernon, M. D. "The Perceptual Process in Reading," The Reading Teacher, XIII (October, 1959), 2-8.

Summarizes evidence from 22 studies relating to nature of the perceptual process in reading and how words are perceived.

94. Vernon, M. D. "I— The Investigation of Reading Problems Today," The British Journal of Educational Psychology, XXX (June, 1960) 146-54.

Presents a critical review of previous research under the following headings: "reading surveys," "techniques of investigating the effects of teaching methods," "studies of groups of retarded readers," "individual studies." (The first paper in a symposium on "The Diagnosis and Remedial Treatment of Reading Difficulties.")

95. Vineyard, Edwin E. and Bailey, Robert B. "Interrelationships of Reading Ability, Listening Skill, Intelligence and Scholastic Achievement," Journal of Developmental Reading, III (Spring, 1960), 174-78.

Bases conclusions on a series of intercorrelations, partial correlations, and second order partial correlation coefficients, using scores made by 114 freshmen college students in tests of reading ability, listening skill and intelligence and grade point averages as measures of scholastic achievement.

96. Weeks, Lewis E., Jr. "Speeding Up Reading: A Self-Help Program for College Freshmen," Journal of Developmental Reading, III (Autumn, 1959), 35-42.

Describes the methods used and the results secured through efforts to stimulate students to improve their speed of reading on their own initiative and effort.

97. Witty, Paul, with the assistance of Ann Coomer and Robert Sizemore. "Individualized Reading — A Summary and Evaluation," Elementary English, XXXVI (October, 1959), 401-12,450.

Reviews opinion and research concering the nature of individualized reading and its efficiency as a mode of instruction.

98. Worth, Walter H. "Promotion vs. Non-Promotion: II, The Edmonton Study," The Alberta Journal of Educational Research, V (September, 1959), 191-203.

"Compares the effect of promotion and non-promotion on the school achievement (including reading) and social personal development" of 66 matched pairs of third and fourth grade pupils "ordinarily categorized as non-achievers."

99. Worth, Walter H. and Shores, J. Harlan, "Does Non-Promotion Improve Achievement in the Language Arts?" Elementary English, XXXVII (January, 1960), 49-52.

Compares the progress in reading of 66 low achievers "who had been non-promoted and repeated the third grade with a like number of low achievers who had been promoted to the fourth grade."

100. Wrolstad, Merald E. "Adult Preferences in Typography: Exploring the Function of Design," Journalism Quarterly, XXXVII (Spring, 1960), 211-23.

Reports the results of an experiment involving 432 adults to determine "the influence of symmetry, of design principles, and of the age, sex, and education of respondents upon their preferences for 'good' or 'poor' typography"(as defined by expert judgment).

101. Zamchick, David. "Paperback Buying Patterns," The English Journal, XLIX (May, 1960), 336-40, 371.

Presents and interprets the results of a one-year survey of purchases from the Teen Age Book Club by students in four eleventh-grade classes (a superior, an average, and two remedial).

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 6, February 1961)

CHILDREN'S LIKES AND DISLIKES ABOUT WHAT THEY READ

J. H. COLEMAN
Executive Analysis Corporation, New York
ANN JUNGEBLUT
Educational Records Bureau, New York

SCHOOL CHILDREN 8 to 12 years of age give substantial evidence that they make discriminating judgments about what they like to read and what they dislike reading.

This has significant implications where it is assumed that children learn more readily and effectively as they find some genuine appeal in what they are led to do or work with. That is, there is evidence to conclude that children's interest ratings of reading materials form a necessary adjunct to adult opinions and statistical indices of readability in order to achieve the optimum grade placement of material for children to read.

How the Children Made Their Ratings

In this investigation, the selection which a child rated was photo-offset on a single $8\frac{1}{2}$ by 13 inch page. The children learned how to do the rating from their teachers who were provided with comprehensive, written instructions.

The rating device is shown in Figure 1. This rating scale was utilized by Thomas N. Jenkins, Ph.D., in developing the vocational interest inventory "How Well Do You Know Your Interests?"

The children were not asked to give their names. It was hoped that such anonymity would encourage them to feel immune from any censure they might think their uninhibited choice would entail. However, teachers were instructed to let the children write their names on the rating sheets if they seemed uncomfortable not doing so.

Better than 99% of the ratings were filled out fully by the children. Only a very few erasures occurred. For this reason, the children seemed to feel fairly sure and secure in the judgments they expressed.

The rating steps were given numerical values from 5 to 0 in descending order from "Like very, very much" to "Dislike very, very much." This was done to compute medians and measures of concentration-dispersion of the ratings.

The Selections the Children Rated

A total of 81 selections, varying in length from approximately 225 to 400 words, were rated by children in Grades 3, 4, 5, and 6. Sixty-three of the selections, taken mainly from children's books published for the trade with a few drawn from textbooks and supplementary texts, were chosen from Books 4, 5, and 6 of Reading for Meaning*. Of the 63, 19 are in Book 4, 24 in Book 5, and 20 in Book 6. Eighteen of the selections were new excerpts from children's books. The content of the selections is illustrated in Table I.

The combined independent ratings of the present authors were obtained on all units in Books 4, 5, and 6. These ratings were judgments of the appeal of the selection to children of the age level involved combined with its appropriateness in terms of educational goals. Each selection was rated on a 3-point scale, namely, Good, Acceptable, or Poor.

Numerical values of 2, 1, and 0, respectively, were assigned to the Suitability Ratings in order to facilitate a comparison of the judgments of adults, experienced as educators, with the children's interest ratings. Only those units from the books with a combined Suitability Rating of 3 or better were utilized in this study. A Suitability Rating of 6 indicates unanimous agreement upon the rating of Good. If two judges rated a selection Good and one Poor, or if two rated it Acceptable and one Good, the combined rating equaled 4, and so forth.

The Lorge Readability Index was used as the primary instrument for the grade placement of the selections. This consists of a standardized formula for analyzing the relative complexity of grammatical structure and the difficulty of the vocabulary of a reading passage. The R.I.'s of the selections employed in this study ranged from 3.1 to 7.0 or, roughly, from a grade level comparable with the beginning of third grade through sixth.

---

* Guiler, W. S., and Coleman, J. H. Reading for Meaning. Chicago: J. B. Lippincott Co., 1955.

FIGURE 1.

Could you read the selection? YES ____ NO ____

IF YOU COULD READ IT, please put a cross (X) in the circle which tells best how you feel about this selection.

○ Like very, very much
○ Like quite a lot
○ Like a little
○ Dislike a little
○ Dislike quite a lot
○ Dislike very, very much

How old were you on your last Birthday? 7 8 9 10 11 12 13 14 15 16 17 18

Draw a circle around the number of the grade you are in 3 4 5 6 7 8 9 10 11 12

Please place a cross (X) after what you are: BOY____ GIRL____

TABLE I

DESCRIPTION OF 24 SELECTIONS FOR FIFTH-GRADERS

No.	Description	N	4	5	6
1.	How Heat Travels: dialogue on thermometers and cooling	177	A	ABE	ABE
2.	Mr. Toad's Tongue: Peter Rabbit sees an unusual tongue	120	A	AB	A
3.	Being a Good Camper: the family puts out a camp fire	86	B	AB	B
4.	The Closest Star: many lifetimes away even by rocket	113	A	AB	A
5.	We Raise Our Hats: courtesy from knights of old	120	A	AB	AB
6.	Where We Get Rubber: from rubber trees to balls and boats	124	B	AB	AB
7.	The Porcupine: how an unusual animal lives	117	B	ABE	B
8.	Peter Drives a Tractor: humorous story about showing off	121	B	ABE	B
9.	Bears on a Railroad Track: a Canadian adventure	168	AE	ABE	AE
10.	Why the Sea is Salty: description and an experiment to do	202	AE	ABE	ABE
11.	Cotton Farming: how cotton grows and is used	104	B	AB	B
12.	The Picnic Basket: a mischievous swan	108	B	AB	B
13.	From Rain to River: description and an experiment to do	114	AE	AB	A
14.	Shore Birds from Everywhere: the penguin and many others	193	AE	ABE	ABE
15.	The School Train: studying in rural Canada	111	B	ABE	B
16.	Shearing the Sheep: a boy helps do a big job	68	B	B	B
17.	The Great Oak: Baltimore orioles build a nest	150	A	AB	ABE
18.	The Pony Express: picked ponies and fearless riders	179	AE	ABE	AE
19.	Plants Feed Themselves: our dependence on plant processes	153	AE	ABE	E
20.	Lichens: how plants change rocks into soil	205	AB	ABE	ABE
21.	Dry and Wet Seasons: a river changes with the seasons	204	AE	AB	AE
22.	Kinds of Trains: moving passengers, freight and circuses	164	AE	ABE	AE
23.	Water Birds: fishing birds and long-distance flyers	176	A	ABE	ABE
24.	Famous Clocks of the World: some large and some amusing	147	AB	ABE	B

N indicates the total number of interest ratings secured; entries under grades 4, 5, and 6 refer to the communities in which children rated the selection.

The Children Who Made the Ratings

Approximately 750 children in grades 4, 5, and 6 in the public schools of Allentown, Penna., Bloomfield, N. J.; and Eastchester, N. Y., each rated 12-14 selections. The first city is situated in an industrial area while the other two may be classified as industrial-residential suburban.

Two elementary schools each in Allentown and Bloomfield, with one in Eastchester, participated in the study, and we wish to acknowledge our heartfelt appreciation to them for assistance in a study which, we hope, has added something of educational benefit. The fact that pupils in two rooms in each grade and school participated made it possible to duplicate ratings under various conditions so as to obtain both intra-school and inter-community samplings of the ratings.

Classrooms varied from 18 to 34 children, averaging about 25. Thus, overcrowded classroom conditions did not seem to obtain. In the 5th grade, there are 41 instances in 3,332 ratings that children could not read the selection. These CNR's involve much fewer than 41 pupils as seen by the 13-year-old 6th grade girl who could not read any of the 5th grade selections offered to her. In most cases, these children read all but one or two selections. Thus, most of these pupils who were unable to read a selection were of normal age for their grade. In two of the participating schools, the project was in the hands of a full-time remedial teacher.

One goal of the study was to have a selection rated by children in seven classrooms, three at the optimal grade level, two in the grade below and two in the grade above. The goal was approximated only for most of the 5th grade selections. Although it was not contemplated that 3rd graders would be included, the interest aroused in two schools elicited a request to permit 35 third graders in two rooms to rate five appropriate selections each during the last month of the school year. The reports available indicate that doing the ratings appealed to the children.

On the assumption that the children might become more discriminating as they acquired practice in the procedure, arrangements were made so that a minimum of ratings on any given unit were among the first any children made.

As with many studies of this type, some unpredictable delays and postponements in participation occurred. This necessitated some juggling of plans so as to insure obtaining a maximum of useful data.

Analysis of Children's Ratings on a Single Selection

The methods utilized in assessing the ratings can be illustrated with the data for selections initially grouped as most appropriate for the 5th grade children.

The distributions of interest-aversion ratings by 202 children on the 20th selection in the 5th grade group are given in Table II. In each grade, ratings are by children in only one classroom per community. The data were obtained from eight classrooms. Three children of normal grade age indicated that they could not read the selection (CNR, Table II).

The spread of the distributions indicates that the children understood how to use the rating method, that they vary in liking or disliking the selection, and that they felt relatively free to indicate their preferences.

How did the children like this selection? It would be ideal if all of them rated it "like very, very much," but they didn't. Hence, their empirical choices require analysis to obtain some of their meaning. The 5th graders' ratings will be considered first. Table III reveals that the median rating for the boys was 4.4, that for the girls 4.2, while their combined median was 4.3. This is interpreted in relation to a value of 4 given a rating of "like quite a lot." In other words, the average rating by these children was definitely on the liking side. The inter-quartile range of 3.4 - 4.9 shows that half the ratings tended to be grouped fairly well around the median of 4.3. It also indicates that 75% of the children rated the selection 3.4 or better. Actually, 81.5% gave a rating showing that they liked the selection to some degree. Furthermore, 20% of the children liked it very, very much.

It is also evident that 12 of the 5th graders disliked the selection, seven of them quite a lot and one very, very much. Is this sufficient reason for discarding the selection for potential use in a remedial reading program for retarded readers in the 5th grade? Further investigation of the data suggests a negative answer to the question. For example, the master tabulation sheet reveals that 11 of the 12 dislike choices were made by children in only one of the three communities. In one of the other communities, 20 of the 22 5th graders rated it in the highest two steps of liking on the scale. It isn't indicated that the children who disliked this selection have a chronic negative attitude. For example, 22 of 23 of them gave the two highest ratings of liking to selection No. 7 in the 5th grade series.

The subject matter of the selection may provide a clue. Written in good style, it discusses in simple terms how lichens change rocks into plant-growing soil. The children who disliked the selection lived in an industrial community. While some of their classmates liked it as much as the children from the other two communities, it may be asked whether those who disliked it either have not acquired as much information about science to feel confident about reading such material as the children who liked it, or may in some other way be less qualified to derive some pleasure from reading material at their level of understanding in the field of science. There is some additional reason to conclude that these children are not quite as sophisticated in some respects as those in the other two communities, as will be seen in a discussion in the next section on the ratings of a relatively humorous selection.

TABLE II

DISTRIBUTION OF INTEREST-AVERSION RATINGS BY 202 CHILDREN IN GRADES 4, 5, AND 6 OF SELECTION 20 IN TABLE I

	Grade 4										Grade 5											Grade 6									
Sex--	Boys (26)				Girls (33)					Both	Boys (38)					Girls (32)				Both	Boys (35)					Girls (41)				Both
Age-- Rating	9	10	11	Total	8	9	10	11	Total	Total	9	10	11	12	Total	9	10	11	Total	Total	11	12	13	14	Total	11	12	13	Total	Total
5	9	2	1	12		7			7	19		5	5		10		1	3	4	14	1				1		1		1	2
4	2	3		5	1	7	3		11	16	1	6	7	1	15		4	11	15	30	1	5			6	6	3		9	15
3	4			4		4	1		5	9		2	4		6		5	2	7	13	10	4		1	15	8	6		14	29
2		1		1		1	1	1	3	4		1	2		3		1		1	4	4	3			7	7	5		12	19
1		1		1		4			4	5		3			3		3	1	4	7	3	1			4	1	1		2	6
0	2		1	3			1		1	4			1		1				-	1	1	1			2	2	1		3	5
CNR						2			2	2								1	1	1										

Grade--	4	5	6		
Allentown	34	30	22		
Bloomfield	25	22	28		
Eastchester	--	18	26		
Total	59	70	76	=	205

TABLE III

READABILITY, SUITABILITY, AND CHILDREN'S RATINGS OF SEVEN FIFTH-GRADE SELECTIONS

Unit	R.I.	S.R.	Grade 4							Grade 5							Grade 6						
			Total N	B N	G N	Boys Mdn.	Girls Mdn.	Both Mdn.	Q_1 - Q_3	Total N	B N	G N	Boys Mdn.	Girls Mdn.	Both Mdn.	Q_1 - Q_3	Total N	B N	G N	Boys Mdn.	Girls Mdn.	Both Mdn.	Q_1 - Q_3
9	4.9	5	48	25	23	3.8	3.9	3.8	3.0-4.9	70	35	35	4.4	4.3	4.4	3.5-5.3	50	23	27	3.6	3.2	3.3	2.4-3.9
10	4.9	5	44	21	23	5.3	4.5	5.1	3.1-5.6	75	43	32	4.6	4.4	4.4	3.6-5.2	83	44	39	3.6	3.7	3.7	2.6-4.2
14	5.0	6	48	21	27	3.8	3.6	3.7	2.6-5.2	66	38	28	4.2	4.1	4.2	3.4-4.9	79	43	36	3.9	3.6	3.8	2.9-4.7
18	5.1	5	50	23	27	5.2	4.9	5.1	4.3-5.6	70	40	30	4.9	4.7	4.8	4.2-5.4	59	33	26	4.8	4.5	4.6	4.0-5.3
20	5.3	6	59	26	33	4.8	4.2	4.4	3.1-5.3	70	38	32	4.4	4.2	4.3	3.4-4.9	76	35	41	3.3	3.2	3.3	2.4-3.9
21	5.3	4	47	23	24	1.5	3.2	3.1	0.7-4.2	43	20	23	4.0	4.3	4.2	3.4-5.1	114	65	49	3.2	3.3	3.2	2.4-3.9
22	5.3	5	43	20	23	5.0	4.1	4.5	3.2-5.3	66	34	32	4.7	4.3	4.5	3.8-5.2	55	31	24	3.8	3.9	3.9	3.2-4.6

To sum up the matter, the structure of the composition in this selection (R.I.-5.3) hardly seems to account for such disliking of it as the negative ratings reveal since children in the same grade in the same schools showed strong or stronger average liking for selections with similar or slightly higher indices of readability. The selection appears to be quite appropriate for remedial reading purposes in the 5th grade except in those schools which may minimize science reading at or before this grade level. The retarded readers might find it difficult to handle, but so long as such a large proportion of their age peers like it, this selection would appear to be one to give retarded readers some stimulus to raise their level of reading ability when it is utilized with a glossary and exercises which are conducive to improving ability to read.

The conclusion just mentioned finds some support from the data for the 4th grade where the combined median rating is 4.4 though the inter-quartile range shows somewhat greater spread of interest and aversion. Two of the 59 children could not read it, of course, but the boys in particular liked it. The selection could be utilized in the 4th grade but seems to be better placed in the 5th, all things considered.

In the 6th grade, the inter-quartile range shows that the ratings are grouped relatively closely around the median 3.3 indicating that the selection is more apt to be liked by children in this grade than not; yet it doesn't appear to have as strong an appeal to children of this age as it does in the next lower grade.

The analysis now turns to the ratings of all the selections originally placed in the 5th grade group.

Interest Ratings for All 5th-Grade-Level Selections

Measures relevant to the appropriate grade placement of all 24 selections originally placed in this 5th grade group were made. Seven of them are presented in Table III. With the medians and inter-quartile ranges will be found the Readability Index, Suitability Rating, and number of ratings (N) for each selection.

The data for selection 18 are interesting. Its medians reveal ratings of liking it on the part of both boys and girls in all three grades--4th, 5th, and 6th. Only 13 of 177 children in seven classrooms from three communities gave it any of the three ratings of disliking. In this sense it may be said to be optimally interesting to children. While the 50 4th graders who rated it did so near the end of the year, the R.I. for this selection is such, 5.1, to expect that more than two children in that grade would have found its structure difficult enough to express more distaste for it. Under the circumstances, the subject matter is of interest. It relates some of the story of "young Johnny Fry" who delivered the first mail from San Francisco to St. Joseph, Mo., on the last leg of the ride by the Pony Express. Along with pertinent details there is drama involving the American dream. "This thing could never be done, they said to each other. Soon a rider came into view. His pony was riding hard and low. As the rider crossed the line, he shouted and threw off the mailbags. Then the rowd went wild. The trip had been made with a few minutes to spare." If all the information we want and need to learn by reading could be presented so appealingly, selecting material for children to read would be markedly simplified.

The interest ratings of selection 6 are curious. Fourth graders like it a lot and so do 6th graders, yet half the 5th graders dislike it, some of them vehemently. This holds for the two classrooms in two different communities, one industrial, one suburban. This would seem logical if either younger or older children exhibited similar variation of choice. The selection consists of a rather factual discussion of both natural and synthetic rubber, though details given have interest, some of it historical. A comparably factual selection about clocks (No. 24) gets a strong interest rating from 5th graders and has the highest R.I. of the group. The sample of 124 children from five different classrooms and three grades in two different communities who rated No. 6 may still be atypical. Or these particular children may simply have been fed up, so to speak, with too much reading about rubber in the 5th grade this year. We don't know. One valid inference seems to be that one never knows until the data are in. Under the circumstances, this selection may be better placed in the 4th grade for remedial reading purposes.

All the facts do not appear in the total grade median of 4.2 so far as selection 14 is concerned. It appears to be interesting enough to appeal to children of average age in the 5th grade who could read it. At the same time, it probably is best used near the close of a series of remedial reading units for that grade despite its R.I. of 5.0 which does not seem to reflect adequately the nature of the concept to be comprehended. The reason is that five of the 193 children to whom it was given could not read it. They were aged nine and ten in the 4th grade, ten and 11 in the 5th grade, and included both boys and girls in each grade. The question is whether these five children, who were able to read most of their selections, could have read this one, and also liked it as well as the other children did, if a glossary of difficult words had been appended to the selection.

Those selections with the highest combined ratings for suitability (6 and 5) revealed interest medians of 4.0 or better, and narrow inter-quartile ranges, with one interesting exception. That is, the children gave relatively high interest ratings to those selections which we rated as most suitable. The exception was selection No. 15. The ratings were: boys, 3.6; girls, 4.2; both, 3.9, with an inter-quartile range of 2.7 - 4.8. The subject matter discusses how children in northern Canada go to school in a railroad coach which arrives at their community once every four weeks and remains one week. The 5th grade children in different communities varied greatly in their attitudes toward this selection. Those from the industrial area rated it 4.3 while those in one of the more suburban communities rated it 2.5. If it may be assumed that the former are relatively less widely read and, for that or other reasons, are more readily interested in subject matter of this kind about other children, this selection would still appear to have an appeal for many retarded readers if they are not too overage for their grade.

The Readability Index of the selections we gave a Suitability Rating of 3 ranged from 4.2 to 5.1. By the R.I., these selections are not too complicated for 5th graders. The children's interest ratings lend some support to the conclusion that selections which appear to be merely acceptable so far as their suitability for educational goals and probable appeal to children is judged, still may be good material for remedial purposes. The reason stems from the fact that the children rated these selections as quite interesting to them. No. 4 is one selection with a Suitability Rating of 3. It has an interest rating median of 4.7. No. 5 is another, with a median rating of 4.2. More than half the 5th graders liked No. 11, as its median of 3.5 reveals. At the same time, 4th graders like it more, so that it appears more appropriate for remedial remedial purposes at that grade level. No. 17, with a median of 4.6 and an inter-quartile range of 3.9 - 5.3, is clearly well placed in the 5th grade in accordance with children's liking for it.

Selection No. 12 with an S.R. of 3, seems to be an exception. Fifth grade girls rate this high for interest, but with the boys the probabilities seem to be 50-50 that they will dislike it. This is an interesting sex difference, for one thing. For another, 95% of the 5th graders in one community rated it "like very, very much," or "like quite a lot." However, the boys and girls in the primarily industrial community varied greatly, some liking it very, very much and some disliking it with equal intensity. This helps to explain the inter-quartile range of 2.0 -5.1. The content of the selection covers a humorous incident about a possibly perfectionist "Aunt" who took her nephew on a picnic and instructed him to watch the picnic basket while she hunted up some fresh drinking water. Wading appealed more to the boy and in his neglect a swan came along, picked up the basket, and started it floating along ahead of him. Auntie saved the day by returning just in time to use her trusty umbrella with the crooked handle, which she always carried to picnics, to retrieve the lunch.

Why does this appeal so much to the children of one community and have such pronounced distaste for some others in another community? It's anybody's guess until further information is obtained, yet one cannot refrain from wondering. Do the children who liked it have more of a sense of humor, or is the atmosphere of the classroom more conducive to expressing humor than with those who have an aversion to this selection? Do the latter feel too strongly that school is such serious business that there is no place for expression of humor in it? Children seem to differ markedly, at any rate, about their feelings toward this selection. Whether it is appropriate for remedial reading purposes may depend, in part at least, upon the sense of humor of those in a position to accept or reject such a selection as desirable or trivial.

With the kind of care that was taken in choosing the selections for rating, it would be surprising to find median ratings for an entire grade group in the "dislike very, very much" or "dislike quite a lot" range. At the same time, a factor to take into consideration is that 21 of the set of 63 units were given S.R. ratings of 3. This may be considered close to a marginal limit of suitability and appeal. It was not anticipated, then, that most of these selections would be rated 3.9 or better in some entire grade, or that only one would have a median below 3.2. This was selection fourteen in the sixth grade series. Even then it may not be the exception which it appears to be since, due to one of the contingencies of carrying out the investigation, it was read and rated only by 62 sixth graders. Their median rating for it was 2.5. Nevertheless, ten of these children rated it either like quite a lot or very, very much.

Summary

An underlying assumption behind this investigation is relevant. It is that school children who are in the process of improving their reading skills are more likely to do so when what they read appeals to them. The object of this study was to see whether children could utilize a six-point scale of liking and aversion to indicate their attitudes or reactions to several prose excerpts apparently within their ability to read and comprehend.

The nature of the distributions of ratings obtained lends support to several conclusions. The indidications are that the children readily learned to utilize the rating scale effectively. The school children who participated in the study appear to have fairly distinct likes and dislikes regarding the selections they were asked to read. They also appear to have made their own, relatively uninhibited choices.

Thus, there are indications that a convenient tool is available to obtain useful evidence about what children enjoy, feel apathetic about, or dislike in what they are called upon to read. It is also indicated that this kind of empirical evidence can help to avoid misjudgments regarding the optimum grade placement of reading subject matter. No selections would be selected solely, of course, on the basis of indices of readability or level of potential reading difficulty. The reason is that value judgments (suitability) of the subject matter would inevitably enter into any selections chosen. In this study it was found that children's interest-aversion ratings of reading subject matter provide an important, even necessary, supplement to a readability index and the pooled judgments of experienced adult educators. This kind of information is highly useful for such purposes as formulating remedial reading programs intended for retarded readers not having serious pathologies, such as vision anomolies or emotional problems, to explain their retarded reading skills.

Several potentially informative areas of investigation come to mind as a result of this study. For example, why do different children react vigorously in opposite directions to the same selection? What relation, if any, does level of intellectual ability have to children's interest-aversion ratings of reading content? Would appending a glossary of the meanings of hard words give children more facility in making interest-aversion ratings? In any case, a method seems to be available for determining which relatively brief reading selections are most liked by the most children.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 6, February 1961)

THE RELATIONSHIPS BETWEEN FIRST GRADE READING READINESS AND SECOND GRADE READING ACHIEVEMNET

MARVIN POWELL
Western Reserve University, Cleveland
KENNETH M. PARSLEY, JR.
Willoughby-Eastlake City Schools, Willoughby, Ohio

THE PURPOSE OF this study was to investigate the relationships between the results of the Lee-Clark Reading Readiness Test administered at the beginning of the first grade and the results of the California Reading test administered to the same students at the beginning of the second grade. The Lee-Clark scores were correlated with the Reading Vocabulary Score, the Reading Comprehension Score, and with the Total Reading Score (the average of the Reading Vocabulary Score and the Reading Comprehension Score) to determine whether the whole test or one of the parts offered the best prediction. A further analysis was done in an attempt to determine whether the reading group placement suggested by the Lee-Clark norms relates to second grade achievement scores.

In a recent preliminary study, the authors found relatively poor relationships between the score on the Lee-Clark Reading Readiness Test and intelligence as measured by the Stanford-Binet. During the course of this earlier investigation the writers became concerned about the adequacy of the Lee-Clark in grouping children in reading sections in the first grade, and particularly with the predictive value of this test in relationship to the reading achievement test scores of these children at the beginning of the second grade. There was some reason to believe that the reading placements suggested by the Lee-Clark norms were not adequate.

A review of the pertinent literature indicated that few studies have been done in this area. Henig (2) related Lee-Clark grade placements with teacher rankings and with grades as a measure of school achievement and found contingency coefficients of .60 and .55 respectively. Moreau (4) related Lee-Clark grade placement with sixth grade achievement grade placement on the California Reading test and found a correlation of .46. Robinson (5) found a correlation of .67 between Lee-Clark and a vocabulary test made up of items taken from beginning reading books. A study by Kerlin (3) related Metropolitan Reading Readiness Test results with Gates Type 3 Paragraph Reading Test finding a correlation of .36 between the two. None of these studies deal with the same material or answer the same questions which concern us.

Sample and Procedure

The Lee-Clark test was given to 863 entering first graders in the Willoughby-Eastlake City Schools in September of 1958. The following year, (September, 1959) these same children were tested at the beginning of the second grade with the Lower Primary edition of the California Reading Test. Only the children who had taken both tests were considered for the final study. Necessarily, those children who were retained in the first grade were also excluded from the study, leaving a final population of 703 children.

Four grade placement scores were recorded for each child: Lee-Clark Score, California Reading Vocabulary Score, California Reading Comprehension Score, and California Total Reading Score (an average of the vocabulary and comprehension scores). The population was then divided into three groups on the basis if the Lee-Clark scores in relation to the interpretation suggested by the California Test Bureau data presented in Table I.

The low group was composed of 49 children whose grade placements were in the range from 0.0 -0.4; the middle group was made up of 365 children with grade placements of 0.5-1.4; and the high group of 297 children was in the 1.5-1.9 grade placement range. The small number of cases in the low group was due to retention of some 45 children who were in the original Lee-Clark low group. Since these children were retained in the first grade there were no achievement test scores available. The above ranges of grade placements were those used by a number of school systems for placing children in appropriate reading groups in the first grade.

The analysis consisted of an attempt to determine the relationships by means of the Pearson product moment correlation technique, Blommer and Lindquist (1). Lee-Clark grade placements were correlated with California Reading test total grade place-

TABLE I

NORMS A-INTERPRETIVE DATA FOR THE LEE-CLARK TEST

Grade Placement Equivalent	Total Score	Expectation of Success
1.5-1.9	56-64	Excellent
1.0-1.4	50-55	Good
0.5-0.9	40-49	Fair
0.1-0.4	32-39	Poor
0.0	31 and below	Very poor

TABLE II

THE SUMS OF SCORES, THE SUMS OF SCORES^2 AND THE CROSS PRODUCTS OF THE VARIOUS GRADE PLACEMENTS FOR THE GROUPS STUDIED

	Low Group (0.0-0.4) N = 49		Middle Group (0,5-1.4) N = 365		High Group (1.5-1.9) N = 297		Total Group (0.0-1.9) N = 703	
	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2
Lee-Clark (W)	9.2	2.9	348.40	371.12	527.2	925.60	878.8	1,298.52
Reading Vocabulary (X)	112.3	265.97	986.4	2,971.92	1,047.9	3,906.73	2,146.6	7,144.62
WX	21.24		996.62		1,849.58		2,858.44	
Reading Comprehension (Y)	78.6	154.28	710.1	1,609.71	764.7	2,154.09	1,553.4	3,918.08
WY	15.35		708.83		1,351.81		2,075.99	
Reading Total (Z)	96.6	206.84	855.9	2,248.73	916.0	2,995.23	1,606.7	5,450.80
WZ	18.99		880.70		1,596.04		2,495.73	

TABLE III

CORRELATION COEFFICIENTS OF THE DATA PRESENTED IN TABLE II

	Low Group (0.0-0.4) N = 49			Middle Group (0.5-1.4) N = 365			High Group (1.5-1.9) N = 297			Total Group (0.0-1.9) N = 703		
	RV	RC	RT	RV	RC	RT	RV	RC	RT	RV	RC	RT
Lee-Clark	.05	.10	.19	.51*	.28*	.66*	.21*	.13*	.72*	.51*	.43*	.82*

* Significant at the .05 level of confidence.

TABLE IV

THE SUMS OF SCORES, THE SUMS OF SCORES^2 AND THE CROSS PRODUCTS OF THE VARIOUS GRADE PLACEMENTS FOR THE GROUPS STUDIED BASED ON WILLOUGHBY-EASTLAKE CUT POINTS

	Low Group (0.0-0.7) N = 139		Middle Group (0.8-1.2) N = 188		High Group (1.3-1.9) N = 376		Total Group (0.0-1.9) N = 703	
	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2	Sum of Scores	Sum of Scores^2
Lee-Clark (W)	63.6	37.08	187.3	191.53	627.9	1,069.91	878.8	1,298.52
Reading Vocabulary (X)	333.2	821.80	523.9	1,613.25	1,289.5	4,709.47	2,146.6	7,144.62
WX	159.08		532.16		2,176.20		2,858.44	
Reading Comprehension (Y)	243.0	493.48	374.9	847	935.5	2,577.55	1.553.4	3,918.08
WY	114.57		379.14		1,582.28		2,075.99	
Reading Total (Z)	291.0	657.58	454.5	1,202.57	1,123.1	3,590.65	1,606.7	5,450.80
WZ	138.25		463.87		1,893.61		2,495.73	

TABLE V

CORRELATION COEFFICIENTS FOR THE DATA PRESENTED IN TABLE IV

	Low Group (0.0-0.7) N = 139			Middle Group (0.8-1.2) N = 188			High Group (1.3-1.9) N = 376			Total Group (0.0-1.9) N = 703		
	RV	RC	RT	RV	RC	RT	RV	RC	RT	RV	RC	RT
Lee-Clark	.26*	.14	.48*	.37*	.25*	.50*	.29*	.27*	.25*	.51*	.43*	.82*

*Significant at the .05 level of confidence.

ments, vocabulary grade placements, and reading comprehension grade placements. This type of analysis was done for the total group and for each of the three sub-groups.

Results and Conclusions

The sums of scores, sums of scores squared and the cross products are presented in Table II. The correlation coefficients are presented in Table III.

An inspection of Table III reveals that the highest degree of relationship exists between the Lee-Clark and the California Total Reading Score for the middle, high, and total groups with correlations of .66, .72, and .82 respectively. The difference between .66 and .72 is not significant. The correlation coefficients for the low group Lee-Clark vs. Total was .19. The correlations between Lee-Clark and Reading Vocabulary for the middle, the high and the total group were .51, .21, and .51 respectively, while the correlations for Reading Comprehension and Lee-Clark for these same groups were .28, .13, and .43 respectively. Only the correlations for the low group were not significant, being .05, .10, and .19 respectively.

In general it would appear that the correlations between the Lee-Clark and the three Reading Scores for the Total Group are more meaningful than the others for general prediction. Specifically, however, the single "best" relationship is between the Lee-Clark and the Reading Total for the Total Group. The failure of the low group relationships to reach significance is probably due primarily to the small sample which does not include children who were retained in the first grade. Had this group been larger, it is possible that the correlations would have been significant, but still not have approximated those of the total group.

The Willoughby-Eastlake schools, where the present study was done, use a cut point of 0.7 and below for selecting children for the low reading group, since this cut point allows more children to be in such a group, than the 0.4 suggested by the California Test Bureau norms. The high group is based on a cut point of 1.3 and above. It was felt that these cut points might be more appropriate than those previously discussed, thus the Lee-Clark grade placements were correlated with the various California Reading Test grade placements for the three groups based on these cut points. The data for the analysis is presented in Tables IV and V.

All correlations except the .14 were significant at the .05 level of confidence but the relationships were generally too low to be useful in prediction. The r of .48 for the Lee-Clark vs. Reading total for the low group in this case was significant while the r for the same comparison in the first breakdown (0.0-0.4) was not. Obviously this increased relationship and the lower r for Lee-Clark vs. total Reading Score for this middle group is due to the redistribution of these groups. Again, however, the greatest degree of relationship is to be found between the Lee-Clark and the Reading Total for the total group.

From the data presented above it seems that the Lee-Clark is useful primarily as a predictor of the Total Reading test results for the entire group. There seems reason to doubt its usefulness in dividing children into reading groups. It would seem that the placement of children into the low group (poor reading prognosis) is hazardous based on Lee-Clark results, since it seems that such placement will be relatively unrelated to their reading achievement at the beginning of the second grade.

The authors feel the Lee-Clark is a useful instrument for predicting the general reading achievement of a total group of first graders. The test is usually used for placing an individual child but there is reason to doubt its adequacy for such precise placement, especially at the lower levels. This is not to say that the instrument cannot be used for initial placement as a crude screening divice, but only to make the teacher aware of the need for her constant re-evaluation of the students. It, therefore, still remains the teachers' responsibility to shift children to the reading groups in which they will function most effectively.

REFERENCES

1. Blommer, Paul and Lindquist, E. F., Statistical Methods in Psychology and Education, Houghton Mifflin Company, Boston, 1960.
2. Henig, Max S., "Predictive Value of a Reading-Readiness Test and of Teachers' Forecasts," Elementary School Journal, 50: 41-6, September, 1949.
3. Karlin, R., "The Prediction of Reading Success from Readiness Tests," Elementary English, 1957, 34, 32-22.
4. Moreau, Margaret C., "Long Term Prediction of Elementary Reading Achievement," California Journal of Educational Research, Vol. 1, No. 4, September, 1950.
5. Robinson, Agnes Soutar, A Study of the Relative Values of Two Standardized Tests in Predicting First-Grade Achievement in Reading, Unpublished Masters' Thesis, Sacramento State College, Sacramento, 1952.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 6, February 1961)

PERSONALITY TRAITS IN READING ABILITY OF WESTERN AUSTRALIAN UNIVERSITY FRESHMEN

A. W. ANDERSON
University of Western Australia, Nedlands

FRESHMEN ADMITTED to the University of Western Australia enroll for professional courses in Arts, Social Sciences, Pure Science, Engineering, Medicine and Applied Science. As part of Orientation Week requirements in 1958, all new entrants were tested on the Cooperative Reading Test (Higher Level Form Y) and the Cattell 16 Personality Factor Questionnaire (Form A). This article reports on investigation into possible relationships between the sub-scale measures on the reading test and the personality factors of the 16 PF test.

The group used consists of 290 males with a mean age of 18.0 years and 125 females with a mean age of 17.8 years. The mean intelligence of this group is 126, standard deviation 7, as measured by the A.C.E.R. B40 Intelligence Test (1).

Table I shows the reading scores for the male and female groups and the combined group. The trend towards lower scores in Speed of Comprehension and in Level of Comprehension is typical of the pattern at the University of Western Australia over the last seven years (2).

Table II shows a brief bipolar description of the personality factors. The general terms are used for convenience, but the fuller analysis of the source trait descriptions should be obtained from the test manual (3).

Table III shows the correlations obtained for the male, female and combined groups. The factors which show a consistent relationship with the reading sub-scales and with Total Reading Score are B, G, I, M, Q1, Q2. All of these show positive correlations except for G, which is negative. Thus in general terms those who score high on the reading test might be described as being more intelligent (B), less conscientious and persevering (G), more sensitive (I), more introverted (M), more radical (Q1) and more self-sufficient (Q2) than those who score low on the reading test.

Apart from the basic factor description, Cattell (4) describes second order factors contained in the 16 PF Questionnaire. Factors M+, Q1+ and Q2+ are components of the Extraversion-Introversion factor and the correlations obtained here describe the Introvert. Factor I+ (and also Factor C-) are contained in Cyclothyme-Schizothyme factor, describing the cyclothyme, although the basic factor A+ shows no correlation with reading score. Factors Ql+ and G- describe the Unbroken Success-Frustration factor, suggesting low frustration associated with reading score.

Table IV shows multiple correlations, beginning with factors I and Q2 combined and adding in successively factors G, Q2 and B. Factor M is not included as it was found that with I included, the effect of M vanishes, the intercorrelation between I and M being +.33. It may be seen that G and Q1 do not increase the multiple correlation as much in Vocabulary and Level of Comprehension as in Speed of Comprehension. Factor B, the intelligence factor, is included last to show the personality factors previously excluding intelligence, in order to suggest the true personality relationships.

Other factors show relationships in specific instances. The scatter for C showed a curved relationship. Among the better readers the correlation became negative, and would be significant for those scoring above the median. Factor F showed a similar trend for Vocabulary. Factor O showed a negative correlation with poorer readers in Comprehension reducing to zero above the median. Q4 shows a relationship in Vocabulary, higher for males who are poorer readers.

Summary

1. The correlations of the scores of 415 freshmen in the University of Western Australia on the Cooperative Reading Test (Form Y) and the 16 Personality Factor Questionnaire (Form A) are shown.
2. Multiple correlations of between .3 and .5 are reported for personality factors and reading.
3. Inclusion of the intelligence measure in the test raises the correlations to between .4 and .6.
4. The general personality description of the good reader, as measured on the tests described, appears to include introversion, emotional sensitivity, self sufficiency, and to alesser extent radicalism and low super-ego strength.

TABLE I

READING SCORES

	Males (290)		Females (125)		All Students (415)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Vocabulary	64.1	8.9	68.4	7.1	65.4	8.6
Speed of Comprehension	60.1	9.1	65.3	10.8	61.6	9.9
Level of Comprehension	61.0	8.7	64.3	8.7	62.0	8.8
Total Reading Score	62.3	8.1	66.5	8.1	63.6	8.3

TABLE II

PERSONALITY FACTORS

	High		Low		High		Low
A	sociable	v	aloof	L	protension	v	relaxed security
B	bright	v	dull	M	introverted	v	practical
C	mature, calm	v	emotional	N	shrewdness	v	naivete
E	dominance	v	submission	O	guilt prone	v	confident
F	surgency	v	desurgency	Q1	radical	v	conservative
G	conscientious	v	casual	Q2	self sufficiency	v	group dependency
H	adventurous	v	shy	Q3	high self sentiment formation	v	poor self sentiment formation
I	sensitive	v	tough	Q4	high ergic tension	v	low ergic tension

TABLE III

CORRELATIONS: 16 PF FORM A AND COOPERATIVE READING FORM Y

Factor	Vocabulary			Speed of Comprehension			Level of Comprehension			Total Reading Score		
	M	F	All	M	F	All	M	F	All	M	F	All
A	10	-11	07	-01	-04	00	01	05	04	-01	-01	-04
B	40*	20^{x}	37*	30*	26^{x}	31*	21*	17	22*	29*	26^{x}	30*
C	01	-15	-05	-08	-18	-14^{x}	-07	-08	-09	-13	-14	-16^{x}
E	06	11	04	05	06	02	03	07	02	-03	15	-02
F	01	-24^{x}	-03	02	-13	01	-02	-07	-01	-06	-17	-06
G	-18^{x}	-17	-14^{x}	-23*	-32*	-22*	-19^{x}	-09	-13	-26*	-18	-20^{x}
H	-03	-12	-05	-05	-07	-06	02	-16	-03	-06	-12	-07
I	31*	22	38*	17^{x}	15	27*	12	12	22*	20*	16	29*
L	10	03	07	03	00	01	03	05	00	03	02	02
M	23*	28^{x}	25*	15^{x}	13	16^{x}	13	15	14^{x}	15^{x}	20	17*
N	15^{x}	-01	02	-05	03	-09	01	-08	-06	-04	05	-09
O	10	03	13	-03	10	06	-22*	01	-11	-03	06	05
Q1	21*	13	15^{x}	22*	07	12	17^{x}	22^{x}	15^{x}	20*	23	16^{x}
Q2	37*	33*	33*	20*	23	19*	23*	27^{x}	24*	26*	29^{x}	25*
Q3	10	-20	03	01	-15	-04	00	-21	-05	00	-15	05
Q4	20*	16	15^{x}	-01	15	05	-06	10	-02	-01	15	04

x significant at 1% level

* significant at 0.1% level

TABLE IV

MULTIPLE CORRELATIONS

	I and Q2	+G	+Q1	+B
Vocabulary	.49	.50	.51	.59
Speed	.32	.38	.39	.46
Level	.33	.34	.36	.40
Total	.37	.41	.43	.49

REFERENCES

1. Australian Council for Educational Research. B40 Test of Adult Intelligence.
2. Anderson, A. W., "Reading Ability and Intelligence of Students," The Educand, 3, 1, November, 1957, pp. 108-112.
3. Handbook for the Sixteen Personality Factor Questionnaire, Institute of Personality and Ability Testing. Illinois, 1957.
4. Cattell, R. B., "Second-Order Personality Factors in the Questionnaire Realm," Journal of Consulting Psychology, 1956, 20, 411-418.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 6, February 1961)

READING IMPROVEMENT IN MILITARY, GOVERNMENT AND BUSINESS AGENCIES

RALPH S. ACKER
U. S. Army Engineer School, Fort Belvoir, Virginia

THE OBJECTIVE of the study was to determine the status of adult reading improvement programs in military, government, and business agencies. *

Prior to preparing the study, increased interest in adult reading ability, demonstrated by professional and lay journals and periodicals, had raised questions concerning the efficiency of adult readers. The typical adult was shown to be performing far below his reading ability potential. In order to develop the reading ability of their employees to a level commensurate with their workload, many MGBA instituted reading improvement programs. It became evident that there was a need for a survey of MGBA adult reading programs, on a national scale, in order to determine how the programs are organized, what their objectives are, what methods, equipment, procedures, and materials are being employed, and what results are being achieved.

Procedure

To determine the status of reading training on an adult level in MGBA, the survey technique of gathering data was used. It was found that those aspects which would have to be surveyed, in order to provide the desired information, grouped themselves into the following areas: general information, programs of instruction, training personnel, test equipment and materials, training equipment - mechanical, training equipment - non-mechanical, and results achieved.

After questions dealing with these seven areas were developed, the adequacy of these items was determined by submitting, for evaluation, a first draft of the survey form to 55 agencies with well established adult reading programs. All of these agencies evaluated the form and submitted suggestions for its improvement. These suggestions were then analyzed and incorporated in the second draft of the survey form. This revised form was then sent to 30 of the 55 agencies in order to determine whether the data received would yield the desired results. At the completion of the preliminary study, certain deficiencies in item structure were found and corrected.

The final form was then sent to 1200 agencies in MGBA. One hundred and seventy-seven of these agencies had reading improvement programs and participated in the study. Approximately 1300 survey forms were mailed and over 800 follow-up letters, telephone calls, and personal contacts were utilized during the progress of this study. It is reasonably certain that all major reading programs in existence in military and government agencies (as of the time of this survey) contributed to this study and there are relatively few, if any, reading programs in major business agencies (as of the time of this survey) that did not contribute to it. The information received from these 177 MGBA was compiled, placed in table form, and analyzed. From these analyses, certain facts, opinions, and trends became evident. For clarity and simplicity, these findings were presented under three separate headings:

1. Findings and interpretations based on facts.
2. Findings and interpretations based on opinion.
3. Trends in adult reading in MGBA.

For brevity reasons, all the findings, interpretations, and trends cannot be presented in this abstract. Only major findings will be included in this brief synopsis.

Findings Based on Facts

1. Programs of Instruction
   a. The most frequently stated objectives of reading programs in MGBA were to increase the speed, comprehension, and vocabulary of adults.
   b. The average program allotted approximately 25 hours to attaining these objectives, although the range was between 19 and 55 hours.
   c. During this time, major emphasis was placed on the mechanics of reading.
   d. Forty-nine percent of the agencies limited attendance to executives, supervisors and selected employees. The remainder set no limitations.
   e. Eighty-two percent of the programs permitted

---

* Further reference to military, government, and business agencies will be indicated by the letters MGBA.

the students to take training on a voluntary basis, most of them offering it during duty time.

2. Training Personnel
    a. Approximately 60% of the agencies operated small reading programs, employing one person to conduct the training.
    b. Almost all the instructors had received some college training. Sixty-seven percent had received some training in the field of education. Of these, only 30% had received training in reading. Of this 30%, less than one-third had received more than nine semester hours of instruction in reading.
    c. Forty-three percent had one year or less of teaching experience in the field of reading.
3. Methods
    a. The most commonly used pieces of mechanical equipment, in order of frequency, were tachistoscopes, reading rate controllers, and training films.
    b. The most commonly used non-mechanical materials, in order of frequency, were reading improvement workbooks, periodicals, and vocabulary workbooks.
    c. Textbooks and workbooks were emphasized by more agencies than any other single type of equipment or material.
    d. The combination of mechanical and non-mechanical equipment and materials was found to be the most effective way to motivate students.
    e. Approximately 60% of the participants employed the group method of instruction, while 30% combined the group with the individual method. The individual method alone was employed by only 10% of the respondents.
4. Testing Devices
    a. Almost all of the MGBA employed at least one standardized test in their programs.
    b. Approximately 64% of them also developed non-standardized tests to meet local needs.
    c. Vision tests were used by one-third of the participants.
    d. Other devices, less frequently used, were intelligence and personality tests.
5. Results Achieved
    a. Approximately 98% of the respondents evaluated the results of their training.
    b. A variety of procedures were utilized, including standardized tests, non-standardized tests, ophthalmograph records, and effects on academic grades.
    c. Standardized test results indicated that reading proficiency improved in all but a few individual cases.
    d. The mean increase in reading speed was approximately 41%, and in comprehension 8%.
    e. The mean percentile increase in vocabulary was approximately 11%.
    f. Ophthalmograph records indicated that the mean number of fixations per 100 words, the mean duration of fixations, and the mean number of regressions per 100 words were reduced, while the mean eye span was increased.

Findings Based on Opinion

1. Programs of Instruction
    a. In order for reading training to be of lasting value, the respondents considered that approximately 20 hours was the minimum length and 31 hours the ideal length of a reading training program.
    b. They considered that these hours should be distributed as follows--no less than three and no more than five hours per week.
    c. In the opinion of the respondents, instructors were the most important factor in making adult reading programs successful.
    d. Respondents were of the opinion that vocabulary and study habits training were helpful in a reading improvement program. Word analysis training was considered to be moderately helpful. Oral reading and literary appreciation training were considered to be of little help in reading improvement.
2. Methods
    a. There was general agreement that the reading rate controller is a highly valuable instrument for adult reading training.
    b. Although they were not considered to be as valuable as the reading rate controller, the tachistoscope and training films were considered to be helpful.
    c. There was overwhelming agreement that reading improvement workbooks were of high value in adult reading improvement programs.
    d. Although vocabulary workbooks and periodicals were considered to be helpful, they were not considered to be as valuable as reading improvement workbooks.
3. Testing Devices
    a. The participants in the study indicated that standardized and non-standardized tests were of definite value in reading training programs.
    b. Opinion was divided regarding the value of the ophthalmograph and vision testing devices, while intelligence and personality tests were considered to be of little value.
4. Results Achieved
    a. Graduates of adult reading programs registered overwhelming satisfaction with the training they received and stated that they would recommend that others take it.
    b. They felt that their reading habits had improved greatly and that they were doing considerably more voluntary reading. They also considered that additional training time would be of value to them.
    c. There was agreement among the supervisors of reading program graduates that the work performance of these graduates improved as a result of their training. They also felt that

the reading habits of graduates improved.

Trends

1. Methods
    a. There is a continuing trend toward group and combination group and individual methods of instruction.
    b. During the three years prior to the completion of this study, there was a decided trend toward the use of non-mechanical materials.
    c. Although the quantity of mechanical equipment in reading programs has not been appreciably reduced, the time allotted to it, per period of instruction, has been reduced.
    d. There is a definite trend toward developing new reading materials in MGBA.
2. Testing Devices

   There was a noticeable trend toward the development of adult reading tests by the respondents.

Overview

1. It is apparent that MGBA feel a definite need for reading improvement programs.
2. The type of program utilized by MGBA is a flexible one, conforming to individual budgets, instructor abilities, or other local requirements. However, the aims are generally the same-- to increase speed, comprehension, and vocabulary.
3. The desire of the participants to improve their reading programs has led to several important changes. The more important of these are:
    a. More of the aspects which produce a mature adult reader are receiving emphasis.
    b. Course lengths are being increased to enable these aspects to be properly developed.
    c. Participants are developing their own reading materials to satisfy local requirements.
    d. Considerable effort is being devoted to the development of testing devices which effectively measure adult reading proficiency.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 MARCH, 1961 NUMBER 7

CONTENTS

Confidential Information and the Guidance Program—Arthur A. Rezny and Ernest Dorow. 243

Art Activities for Mentally Handicapped Children—Richard G. Wiggin. 251

A Study of the Socio-Economic Backgrounds of Parents of Miami University (Ohio) Students—Joseph C. Seibert and C. Neale Bogner 259

An Experiment with the Controlled Reader—Forbes Bottomly 265

A Beginning Reading Vocabulary—Daniel W. Fullmer and Clifford J. Kolson 270

The Relationship Between Science Proficiency and Training Among Liberal Arts Graduates—Jacob Tuckman and William Etkin 273

The Effect of Training in Listening for Specific Purposes—Sue E. Trivette 276

Research News—John Schmid 278

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

CONFIDENTIAL INFORMATION AND THE GUIDANCE PROGRAM

ARTHUR A. REZNY
University of Wisconsin—Milwaukee
ERNEST DOROW
West Allis Public Schools, West Allis, Wisconsin

THE PLACE OF confidential information in guidance programs has been the source of much disagreement among educators and other professional people. The guidance program of any school system can only function to its fullest with a basic understanding of the job to be done and the mode of operation to accomplish the assignment. The authors were confronted with a very simple problem. What is meant by confidential information and how should it be used? However, to define confidential information is a relatively simple matter. For a complete staff to understand confidential information is another problem.

Significance of the Problem

This problem was magnified by the absence of a discussion of confidential information in current literature in guidance publications. The authors felt obligated to discover what the experts, administrators, and teachers thought confidential information was and from such an understanding, how the total guidance program could function beneficially for the school child.

It has been most unfortunate that most of the books on guidance have not been written from a philosophical background of education and, in particular, public education.

In order to clarify objectives and to give continued impetus in one direction, guide posts which point to education in the public schools and which lead toward clarification and the culmination of an idealistic guidance program are as follows: 1) A democratic philosophy must be inherent in the school system and must be in operation in all situations. 2) Instruction is supreme. This is the function of the school. Instruction is operative and creative. Teaching is the activity by which the instructional program (curriculum) is placed into operation by the teacher through the process of guiding and directing individual learning within or without the classroom. 3) The public educational system must be non-sectarian, non-partisan, classless, and impartial. 4) The school is an extension of the home. 5) The school is only one of the social agencies in a community responsible for the total life of the child or adult.

Although there are other important guide posts in a philosophy of education in a democracy, it is of utmost importance that the above principles are followed if a complete and functional guidance program is desired.

Definitions of Terms

The concept of confidential information must operate within the guide posts. Definitions sometimes create problems. Sharp discriminations have frequently resulted from differences in definitions, partly because of the nature of definitions. Definitions, according to Lundberg (1), do not reveal what the data in question "really are." In defining information as complicated as that with which we are dealing, with so many variables involved, one is bound to run into disagreement over what is more or what is less. Yet, definitions are necessary for communication. We must remember simply that they are "constructs of convenience."

Because no available source material in the field of guidance literature used in this study quoted or definitely defined confidential information, the dictionaries of Webster and Thorndike were consulted.

Thorndike defines confidential as: "adj. 1. told or written as a secret. 2. showing confidence. 3. trusted with secrets, private affairs, etc..." Information is defined as: "n. 1. knowledge, facts, news."

Webster defines confidential as: "adj. 1. communicated in confidence, secret. 2. indicating intimacy or communication in confidence. Enjoying or treated with confidence." Information is defined as: "n. 1. act or process of informing; specific; communication of knowledge or intelligence; also accusation. Knowledge derived from reading, observation, or instruction; esp. unorganized or unrelated facts or data."

In view of the above definitions and in view of the setting of a pupil in a school system, it was the objective of this study to discover the interpretation of confidential information by various colleagues in the field of education.

Procedure

The opinionnaire of six questions stated singularly in Tables II-VII was formulated and sent to 101 sources of information.

1. Thirty authorities in the field of guidance were sent opinionnaires. (Authors of the latest editions of new textbooks in guidance were consulted. Broad areas of land sections in this group were attempted, the East, West, North and South).
2. This same section coverage was followed in sending the opinionnaire to guidance and personnel officials of leading universities and colleges in the United States. This total was 42 in number.
3. The opinionnaire was sent to guidance and personnel officials in the high schools in the state of Wisconsin. Twenty-eight schools were selected on a geographical basis. The total included ten high schools in the Milwaukee metropolitan area.
4. A large metropolitan guidance clinic was contacted.

Table I gives a categorical and total pollee response to the opinionnaire. According to Powell, an authority on public opinions, the 44 percent total response is a very good return (2). It is interesting to note that the greatest response came from the authorities in the field.

An analysis of the answers to the request for a definition or analysis of confidential information and the five related questions on the opinionnaire follow with interpretations and conclusions based upon the recorded statistical data. A summary of conclusions based on the available data appears at the end of the study.

Interpretation of Data-Table II, Question No. 1

1. An area called confidential information does exist in the field of guidance and that varied points of view exist as concerns this area. Of all sources contacted, 84 percent formulated an opinion concerning their definition of confidential information, two percent felt that no information was confidential, 29 percent felt that certain personal information such as wrong-doing, family problems, behavior situations, etc. should be classified as confidential information, 23 percent felt that the criteria of judgment concerning whether information was confidential depended upon the professional attitude of those who had access to such information, and 18 percent felt that all information given to the counselor or teacher should be classified as confidential.
2. Of the three groups contacted, the statistical data varies considerably within each group. Regarding the thought that information given to a counselor in confidence should be classified as confidential is upheld by five percent of the recognized authorities, 54 percent of the guidance personnel at the college and university level, and 0 percent of the guidance personnel in high schools.
3. The variations in responses seem to be based on individual opinion and experience of the pollees rather than from a basic understanding as to the purpose of a guidance program in a public school system thus leading to variations in the individuals definition of confidential information. As a result of this thinking, it would appear that the determiner as to whether any information is confidential or not depends upon the information itself and the person who makes the interpretation.

Interpretation of Data-Table III, Question No. 2

1. A diversity of opinion exists at all levels concerning what information should or should not be contained in the cumulative records. Thirty-nine percent of the pollees felt that the cumulative records should be all-inclusive as concerns information involving the student; 50 percent of the pollees felt that certain restrictions are necessary; 11 percent failed to answer, indicating either no opinion or interest in the question.
2. Comparative data indicates that proportionately within the polled groups, they are rather divided in their opinions as concerns the question of inclusiveness of cumulative records. Thirty percent of the recognized authorities in the field of guidance felt that the records should be all-inclusive, 38 percent of the college guidance personnel felt likewise, while 55 percent of the high school guidance teacher personnel strongly favor this idea. All pollees of the latter group responded to this question while the recognized authorities and college guidance personnel groups responded with a 20 percent and eight percent reply respectively in the "No Answer" category. The 55 percent response of the high school personnel group supported this percentage with a mass opinion that much information is not available to them. Fifty percent of the recognized authorities recommended that restrictions govern the information contained in the cumulative record. This group emphasized their responses by steadfastly suggesting that many of our faculty members are not mature enough to handle all personal information about students and secondly, suggesting that we, as a professional group, must realize that inadvertently or otherwise, ethics is a forgotten word when student information is involved.
3. All groups were guided in answering this question not by the nature and content of the information involved, but rather by the professional attitude of those who use or who have access to this information. This implies, by inference, that teachers are many times the cause of this attitude. If one of our guide posts in education is that instruction is supreme, and if one adheres to the principles of democracy which entail the individual as of inestimable value, and that his development is the sole objective of society, can we restrict information when professional people are concerned with the child's development?

TABLE I

A BREAKDOWN OF THE CATEGORICAL RESPONSES

Categories	Total	Responses	Categorical Percentage	Total Percentage
Authorities consulted	30	20	67	20
Directors of Guidance and Personnel Services--Colleges and Universities	42	13	31	13
Directors of Guidance and Personnel Services--High Schools in Wisconsin	28	11	39	11
Professional Clinic Agency	1	0	0	0
Total	101	44	137	44

TABLE II

RESPONSES TO QUESTION NO. 1
(The question: What is your definition or analysis of information that might be classified as confidential as concerns the individual cumulative record?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
All information confidential	2	10	2	15	2	18	6	14
Unusual personal behavior situations, wrongdoing, illness, or private family data	7	35	1	8	5	46	13	29
That information given in confidence to counselor	1	5	7	54	0	0	8	18
Only objective information that a faculty is competent to use	6	30	2	15	2	18	10	23
No answer	4	20	1	8	1	9	6	14
No information confidential	0	0	0	0	1	9	1	2
Totals	20	100	13	100	11	100	44	100

Interpretation of Data-Table IV, Question No. 3

1. The data again reveals a diversity of opinions among and within the groups concerning the availability of student information to all interested professional personnel. Forty-three percent of the pollees believed that the classroom teacher should have access to all information while 48 percent responded negatively concerning this question.

2. The recognized authorities responded with a 55 percent "no" response accentuating thoughts expressed in many of their texts which tend to focalize the danger of misinterpretation of factual material when student information is forwarded from one teacher to another and from indiscriminate usage of this material by teachers at other than professional level.

3. From the large 62 percent "no" responses recorded by the college and university group, three attitudes apparently supported this response. a) Both the recognized authorities and the college-university groups have and do emphasize the role of the specialist as concerns the interpretation of informational data or b) that these two groups realize that teachers and many other professional personnel are inadequately prepared to interpret data and student records and to utilize this material in planning each student's program to solve individual problems within the classroom environment; c) this also suggests a probable revision in our teacher training program as concerns curricula or an extension of the present four-year program.

4. The high school guidance personnel, namely classroom teachers, responded with an 82 percent response to this question and indicated by word opinion that there is an urgent need for the classroom teacher to have all available information if meeting and solving individual needs within the classroom environment is to be accomplished.

5. The data reveals that a lack of continuity in philosophy concerning this question exists between high school guidance personnel and the college-university groups.

6. It was also apparent that the specialists in the field of guidance felt that many classroom teachers are not qualified to interpret and to have access to all information including the data described as confidential information. Unethical practices concerning the use of this information in terms of lunchroom and daily conversation rather than professional application of such material by teachers was reported as a significant reason for a 20 percent "no" response to this question by the recognized authority groups.

Interpretation of Data-Table V, Question No. 4

1. Many of the pollees felt that a code of ethics would restrict dissemination and communication of student information. Twenty-three percent felt that while a code of ethics is necessary and appropriate, it would result in eliminating unprofessional practices among teachers and professional personnel, yet not result in the withholding of student information if and when necessary.

2. Statistics reveal that a large portion of the polled personnel in all groups (36 percent) are not familiar with the proposed code of ethics and responded in kind.

3. High school guidance and teacher personnel, basing their opinions upon past experiences, voiced a 73 percent "yes" vote to this question maintaining that codes for the most part tend to prevent the communication of information vital and necessary to the indicating of individual differences and the preparation for educational planning at the classroom level. Comments also indicated a reluctance to accept many of the so-called personnel classified as specialists in the field of guidance.

4. College guidance personnel and the recognized authorities in the field of guidance who answered the question were equally divided in their responses.

5. The recognized authority group tallied the highest (46 percent) in the "no answer" group, indicating unfamiliarity or lack of interest in the question.

Interpretation of Data-Table VI, Question No. 5

1. This question caused a greater range of response than any other question in the opinionnaire, indicating that a great deal of uncertainty exists as concerns the legal aspect involving communicated information from student to teacher or counselor. Eighteen percent were positive no legal obligation was involved. Seven percent were doubtful, 30 percent did not know, 20 percent failed to answer, five percent thought some legal obligation was involved, while two percent suggested state statutes might differ from state to state on this question.

2. The recognized authorities in the field of guidance appeared to be more positive than other categorical groups in that 25 percent felt that no legal obligation was involved. However, 35 percent of this group responded to the "I don't know" response while 25 percent gave a "no answer" response. This total of 60 percent, indicating doubt or uncertainty by the recognized authorities, possibly explains the lack of source material concerning this question.

3. The college and university group (38 percent) and the high school group (27 percent) felt that a legal obligation was involved while the recognized authorities were positive there was no legal obligation involved and responded with a 0 percent. Thus, there appears to be a wide breach of opinion between this group who are the authors of the texts, brochures, et cetera, and those in the field at the high school and college level.

4. A great deal of doubt exists in the minds of all pollees as concerns the legal aspects of this question and to the extent and limitation governing the field of guidance and the possible legal ramifications.

TABLE III

RESPONSES TO QUESTION NO. 2

(The question: Should the cumulative records be all inclusive or should limitations or restrictions govern the information contained therein?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
All Inclusive	6	30	5	38	6	55	17	39
Restricted	10	50	7	54	5	45	22	50
No Answer	4	20	1	8	0	0	5	11
Totals	20	100	13	100	11	100	44	100

TABLE IV

RESPONSES TO QUESTION NO. 3

(The question: Should the classroom teachers, as well as other interested professional individuals, have access to all information available involving individual students, if and when necessary?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
Yes	5	25	5	38	9	82	19	43
No	11	55	8	62	2	18	21	48
No Answer	4	20	0	0	0	0	4	9
Totals	20	100	13	100	11	100	44	100

TABLE V

RESPONSES TO QUESTION NO. 4

(The question: Do you believe that a code of ethics such as proposed by the N. E. A. restricts the communication and dissemination of information to all teachers interested in a particular student's problems?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
Yes	6	30	4	31	8	73	18	41
No	6	30	3	23	1	9	10	23
No Answer	8	40	6	46	2	18	16	36
Totals	20	100	13	100	11	100	44	100

TABLE VI

RESPONSES TO QUESTION NO. 5

(The question: What legal obligation, if any, is involved as concerns personal information communicated by a student to a counselor or teacher under oath or otherwise?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
None	5	25	2	15.4	1	9.2	8	18
Doubtful	0	0	2	15.4	1	9.2	3	7
Some	2	10	0	0.0	0	0.0	2	5
I don't know	7	35	2	15.4	4	36.3	13	30
Differ by states	1	5	0	0.0	0	0.0	1	2
No answer	5	25	2	15.4	2	18.1	9	20
Yes	0	0	5	38.4	3	27.2	8	18
Totals	20	100	13	100.0	11	100.0	44	100

TABLE VII

RESPONSES TO QUESTION NO. 6

(The question: Can a personal cumulative record be the source of legal action by a student against any member or members of an educational institution?)

Responses	Recognized Authorities		Guidance and Personnel Directors, Col. and Uni.		Guidance and Personnel Directors, High School		Total	
	No.	%	No.	%	No.	%	No.	%
I don't know	7	35	1	8	5	45.0	13	29.5
Yes	1	5	5	38	4	36.1	10	23.0
No	0	0	0	0	1	9.9	1	2.0
Doubtful	7	35	6	46	0	0.0	13	29.5
No Answer	5	25	1	8	1	9.9	7	16.0
Totals	20	100	13	100	11	100.0	44	100.0

5. No group, as a whole, of those polled, was positive or could quote source material that could help resolve the question.

Interpretation of Data-Table VII, Question No. 6

1. The responses tend to fortify the conclusions drawn from question number five and again indicate the remarkable degree of uncertainty that exists concerning the legal aspects involved in this phase of guidance.

2. The most interesting fact concerning the statistical data in this question is the change of response to this question as compared to the previous question, namely: a) in the authority category, there were 0 percent doubtful cases concerning the legal aspect of question number five, while in question number six, which also involves a legal interpretation, 35 percent express a doubtful response; b) in question number five, no authority responded with a "yes" answer while in question number six, five percent felt a "yes" response was necessary; c) 25 percent of the authorities were emphatic in responding "no" to the legal interpretation of this question, while in question number six, again involving a legal interpretation, a 0 percent of "no" responses was recorded. This marked degree of uncertainty among those who help formulate guidance programs, by virtue of the many forms of literature used as source material by guidance personnel at all levels, reflects this influence by these consumers.

3. The same phenomena of responses to this question, as expressed by the recognized authority group in questions number five and six, are comparable in the other two categories, also indicating the marked degree of uncertainty concerning the legal implications regarding the oral or the written word.

4. Many guidance personnel at all levels misconstrue the fact that all written guidance records are public records. Authorities disagree as to whether these records can be legally subpoenaed by law, and be used as evidence.

Summary

While this investigation is probably limited in its scope, the data appears to be significant and valid. Some major conclusions based on the available data are as follows:

1. That an aspect of guidance termed or described as confidential information does exist in the field of educational guidance.

2. No general rule exists as concerns the description or definition of confidential information in the field of guidance.

3. Many points of view concerning confidential information exist at all levels in the educational field.

4. The common yardstick for measuring who should have access to confidential information is based on personal opinion rather than on a prevailing professional attitude among all teachers.

5. That apparently, as a profession, many within the teaching profession feel that teachers do not possess the qualifications identified with a profession.

6. That a code of ethics is necessary, when dealing with confidential information.

7. That the higher echelons of education feel that guidance and counseling is a separate field that requires highly specialized training.

8. It appears that our present college curriculum needs revision as concerns future training for teachers... if we are to accept the premise that education and guidance are inseparable and that the teacher is the hub of any functional guidance program.

9. The search of legal references has not produced any evidence that guidance counseling comes within the scope of activities to which professional immunity is granted.

10. Guidance and counseling qualifies as a profession in its educational requirements and in its code of ethics.

11. The teaching profession is at present suffering from a lack of statutory immunity such as is granted to other professions which make use of highly personal information.

12. The teacher-counselor is ethically bound by his professional philosophy, if not by the legal statutes of all states, to consider the counselee's activities, his failure to adjust and his particular problems as professional secrets.

13. Teacher-counselors should know under what conditions and to whom, and only to whom, they may make use of privileged communication without being in danger of legal action for libel or slander.

14. While it may be reasonable to assume that the general principles of statutory law dealing with privileged communication will be applied to counseling if a case arises that is not within the letter but is plainly within the spirit of the law, it is nevertheless contended that guidance counseling is of such a nature that it is in the best interest of the counselor, the counselee, and the public at large that the profession be granted statutory immunity.

REFERENCES

1. Simpson, George E. and Yinger, J. Milton, Racial and Cultural Minorities, New York: Harper and Brothers, 1953.
2. Powell, Norman John, Anatomy of Public Opinion, New York: Prentice-Hall, 1951.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

ART ACTIVITIES FOR MENTALLY HANDICAPPED CHILDREN *

RICHARD G. WIGGIN
Arlington County Public Schools, Arlington, Virginia

FOR SEVERAL YEARS, individual Arlington art teachers have been teaching one or two so-called “special” art classes in their daily teaching schedules. These classes normally consisted of from ten to 25 students of junior or senior high school age, who were grouped together principally because their educational needs appeared to call for special help. More recently, they have been defined as “mentally handicapped,” and have been classified into two sub-groups: educable retarded, 55-75 I. Q. and slow learners, 75-90 I. Q.

Recently, the number of such special art classes has increased. This last year, 1957-1958, three Arlington art teachers were each given one or two such classes. As a result, the three art teachers and the art supervisor met informally in the fall of 1957 mainly to discuss common problems they faced in teaching these classes. Out of this discussion, three questions were developed. It was felt that answers to these questions would materially help to resolve the common problems discussed.

1. What art activities are popular with mentally handicapped children of high school age in our schools?
2. Why are these activities popular?
3. Can a “yardstick” be developed which will predict the degree of popularity of a newly-introduced activity?

Design of the Study

To find answers to the above questions, the following steps were taken:

1. A list of “most popular” art activities, individually derived by each of the three art teachers, was compiled.
2. A master list, combining the three lists by the three art teachers, was compiled. (This list is reproduced in Table I.)
3. The recurring characteristics common to these activities were sought. (See Table II for 41 characteristics.)
4. The major factors or principles which appear to be basic to these characteristics were extracted.
5. A “yardstick” to rate the relative popularity of newly adopted art activities, using the above characteristics as a key”, was developed.

Results[2]

After compiling the master list of 36 popular art activities (Table I), 41 characteristics were found to be common to some or all of the above activities (Table II). It was further found that 32 of the 41 characteristics were inter-correlated below the 5% level of significance and could be considered a “normative factor,” since they were found to be common to 24 or more of the 36 popular art activities (Table III). The remaining nine characteristics found in less than 24 of the activities proved also to be inter-correlated with each other below the 5% level of significance and, thus, could be considered a “differentiating factor” (Table IV). This then became the “yardstick” for which we were looking.

As a final step, the 36 popular art activities were then placed in descending order of popularity, according to the degree to which they contained the positive poles of the nine characteristics present in the “differentiating factor,” or “yardstick” (Table V).

Conclusions and Next Steps

Tentative conclusions from the first phase of our study of the art preferences of mentally handicapped children are outlined below:

A. There were 51 art activities which we found to be popular with mentally handicapped children of high school age in Arlington County (reproduced in Table I).

B. The reasons for the popularity of these art activities with mentally handicapped children lies, at least partly, in the fact that they contain 32 general characteristics (reproduced in Table III).

C. The more popular a specific art activity, the more likely it is to contain the following characteristics:

1. Material has a definite weight

---

* Footnotes will be found at the end of the article.

TABLE I

POPULAR ART ACTIVITIES FOR MENTALLY HANDICAPPED CHILDREN*

1. Plaster Carving (abstract, concrete)
2. Linoleum Cutting and Printing
3. Copper Repousse
4. Copper Enameling--Jewelry
5. Copper Wire Jewelry
6. Raised Copper Trays and Bowls
7. Finger Painting
8. Painting (outside landscape)
9. Figure Drawing (contour)
10. Clay Sculpture
11. Paper Sculpture (abstract, concrete)
12. Weaving
13. Mobiles and Stabiles
14. Water Color Painting (fruits and vegetables)
15. Papier Mache (of people, animals)
16. Chalk Drawings (of fall leaves)
17. Clay Pottery
18. Printing (scrap, string, spatter, potato)
19. Diorama
20. Basketry
21. Hooking
22. Crayon Techniques (melted, resist, etching, side, imprint)
23. Cut Paper
24. Stuffed Rag Animals (girls only)
25. Puppets and Marionettes
26. Embroidery
27. Crocheting
28. Knitting
29. Leather Tooling
30. Scratch Board
31. Nature Study (paint, construction paper, papier mache, wire)
32. Salt and Flour Plackets
33. Wood Carving
34. Plaster Sculpture
35. Paper Mosaic
36. Self-Portraits
37. Wood Burning
38. Chenille Craft
39. Rattan Weaving
40. Book Ends (with etched aluminum faces)
41. Applique (felt, wood)
42. Wire Gadgets
43. Silk Screen
44. Small Wood Crafts
45. Simple Electrical Wiring and Wiring Repair
46. Framed String Design
47. Limited Mosaics (embedding tesserae into a 3 D form)
48. Design of Games (toss games, puzzles, space game 3 D)
49. Mobiles Made from a Pattern (wood or metal material)
50. Plant Flowers
51. Papier Mache Grotesque Masks

---

* Number of different popular art activities reported by Anderson, Blake, and Boulware. Phase 1. November 5, 1957 (including ten additional items 42-51). March 28, 1958.

TABLE II

CHARACTERISTICS FOUND IN THE 36 POPULAR ART ACTIVITIES

Positive Pole	Negative Pole
Tools and Materials	Tools and Materials
1. Material offers definite resistance	1. Material offers no resistance
2. Material is shaped in process of use	2. Material is not shaped in process of use
3. Material has definite weight	3. Material is weightless
4. Material is permanent, solid, hard	4. Material is impermanent, fragile, soft
5. Manual arts tools used	5. Fine arts tools used
6. Acceptable raw material having monetary value	6. Unacceptable raw material without monetary value
7. Unworked material has structure	7. Unworked material lacks structure
8. Tools and materials individually controlled	8. Tools and materials shared in common
9. Design carried out with a mechanical aid	9. Design carried out without mechanical aid
10. Individual tool storage, fixed and unchanging	10. Group tool storage, flexible, changing
Objectives	Objectives
11. Personal adornment	11. Not for personal adornment
12. Utilitarian	12. Esthetic--decorative
13. End product acceptable to adults	13. End product not acceptable to adults
14. Product designed as a gift	14. Product expendable
Work Process	Work Process
15. Larger arm movements predominant	15. Small finger movements predominant
16. Individual sequence, and varied	16. Step by step sequence
17. Skills develop during execution	17. Little chance for skill development during execution
18. Tactile sensations dominant	18. Tactile sensations weak or absent
19. Product designed from something seen	19. Product designed from something imagined
20. Repetitious, rhythmic, predictable results	20. Variable, unrhythmic, unpredictable results
21. Sensory reactions strong	21. Emotional reactions strong
22. In and out movements	22. Movements on one plane
23. Squeezing, hammering, gouging, cutting, pushing (vigorous muscular action)	23. Finger manipulation only
24. Preference for independent activity	24. Preference for group activity
25. Strongly competitive activity	25. Activity lacks competitiveness
26. Rolling and stretching activity	26. No rolling and stretching activity
27. Daily product growth visible	27. Daily product growth not visible
Teaching Process	Teaching Process
28. Show finished product, outline steps singly	28. Outline steps together
29. Teacher directions introducing each step	29. Teacher directions during introductory stage only
30. Present restricted choices, with obvious differences	30. Present wide choices involving subtle differences
31. Give positive directions such as "do this"	31. Give negative directions such as "don't do this"
32. Warn child before questioning	32. Question child without warning
33. Teacher reflects calmness	33. Teacher reflects tension
34. Students free to move within work area	34. Students confined to one space for work
35. Teacher adjusts daily activity to varying emotional moods of students	35. Teacher does not adjust activity to varying emotional moods of students
Appearance	Appearance
36. Color dominant	36. Color subordinate or absent
37. Relief	37. Flat
38. Controlled classroom organization	38. Spontaneous classroom organization
39. Symmetrical	39. Asymmetrical
40. Definite form	40. Indefinite form
41. Geometric shapes dominate	41. Amorphous shapes dominate

TABLE III

CHARACTERISTICS CONSTITUTING THE NORMATIVE FACTOR (POSITIVE POLES ONLY)

Tools and Materials

1. Material offers definite resistance
2. Material is shaped in process of use
3. Unworked material has structure
4. Tools and materials individually controlled
5. Design carried out with a mechanical aid
6. Individual tool storage, fixed and unchanging

Objectives

7. Personal adornment
8. Utilitarian
9. End product acceptable to adults

Work Process

10. Larger arm movements predominant
11. Individual sequence, and varied
12. Skills develop during execution
13. Tactile sensations dominant
14. Product designed from something seen
15. Repetitious, rhythmic, predictable results
16. In and out movements
17. Squeezing, hammering, gouging, cutting, pushing (vigorous muscular action)

Work Process (Continued)

18. Preference for independent activity
19. Strongly competitive activity
20. Rolling and stretching activity

Teaching Process

21. Teacher directions introducing each step
22. Present restricted choices, with obvious differences
23. Give positive directions such as "do this"
24. Warn child before questioning
25. Teacher reflects calmness
26. Students free to move within work area
27. Teacher adjusts daily activity to varying emotional moods of students

Appearance

28. Color dominant
29. Controlled classroom organization
30. Symmetrical
31. Definite form
32. Geometric shapes dominate

TABLE IV

THE DIFFERENTIATING FACTOR*

Positive Pole	Negative Pole
1. Material has a definite weight	1. Material is weightless
2. Material is permanent, solid, hard	2. Material is impermanent, fragile, soft
3. Manual arts tools are used	3. Fine arts tools are used
4. Relief	4. Flat
5. Show finished product, outline steps singly	5. Outline steps together
6. Acceptable raw material having monetary value	6. Unacceptable raw material without monetary value
7. Sensory reactions strong	7. Emotional reactions strong
8. Product designed as a gift	8. Product expendable
9. Daily product growth visible	9. Daily product growth not visible

* Showing the positive and negative poles of the nine characteristics found in less than 24 of the 36 popular art activities.

TABLE V

RELATIVE POPULARITY OF THE 36 ART ACTIVITIES*

Rank Order	Activity	Characteristics** in Factor II
1	Raised copper trays and bowls Leather tooling	9
2	Linoleum cutting and printing Copper repousse Copper enameling--jewelry Weaving Clay pottery Basketry	7
3	Copper wire jewelry Wood carving	5
4	Papier mache (of people, animals) Hooking Stuffed rag animals (girls only) Puppets and marionettes Plaster sculpture	3
5	Plaster carving (abstract, concrete) Clay sculpture Salt and flour plackets	1
6	Embroidery Crocheting Knitting	-1
7	Mobiles and stabiles	-3
8	Paper sculpture (abstract, concrete) Printing (scrap, string, spatter, potato) Paper mosaic	-5
9	Crayon techniques (melted, resist, etching, side, imprint) Scratch board	-7
10	Diorama	-8
11	Finger painting Painting (outside landscapes) Figure drawing (contour) Water color painting (of fruits and vegetables) Chalk drawings (of fall leaves) Cut paper, nature study, and self-portraits	-9

---

* Showing the number of constructs from Factor II found in each of the 36 popular art activities. The activities are ranked in order according to popularity, using a "key" derived from Finger Painting, which the three art teachers voted the "least popular" of the 36 activities.

** The score in this column represents the algebraic sum of the poles of the nine characteristics found at each rank order. Thus, a score of 9 opposite rank 1 means that the positive poles of all nine characteristics were present. A score of -9 opposite rank 11 means that the negative poles of all nine characteristics were present. A score of 5 opposite rank 3 means that 7 positive, 2 negative poles were present.

2. Material is permanent, solid, hard
3. Manual art tools are used
4. Relief
5. Show finished product, outline steps singly
6. Acceptable raw material having monetary value
7. Sensory reactions strong
8. Product designed as a gift
9. Daily product growth visible

D. The less popular an art activity, the more likely it is to contain the opposing poles of the above nine characteristics. These opposing poles are:

1. Material is weightless
2. Material is impermanent, fragile, soft
3. Fine arts tools are used
4. Flat
5. Outline steps together
6. Unacceptable raw material without monetary value
7. Emotional reactions strong
8. Product expendable
9. Daily product growth not visible

E. It appears possible to develop a "yardstick" against which any new art activity can be measured, and its relative degree of popularity predicted. This yardstick should contain the nine characteristics found in Factor 2, each described by its two opposing poles. (The validity of this last statement will be examined in the second phase of this study to be conducted in the 1958-1959 school year.)

F. Apparent ramifications: From a study of the 41 characteristics of art activities popular with mentally handicapped children, the following opinions seem apparent and warrant further study and validation:

1. A mentally handicapped child's definition of reality seems to be built around a physical reality which has permanence, solidity, and three-dimensional depth.
2. In addition, objects seem most real to the mentally handicapped child when they evoke strong tactile reactions and involve large muscle activity of a strenuous nature.
3. His definition of reality is further delineated and conditioned by the attitudes and preferences of his adult peers. Their preferences are for products which apparently fill all of the above physical conditions, plus having monetary value, and meeting other less defined adult standards of craftsmanship.
4. The use of the project should preferably be as a gift.
5. Tools used should be oriented towards those normally used in the industrial arts field.
6. The teaching process apparently involves several specific characteristics:
    a. The principal objective of the teaching process should be to maintain and enhance the individual student's sense of security. This may even involve turning over a whole class period to the therapeutic "talking out" of personal problems.
    b. In the introductory phase of a new lesson, sample finished products should first be shown, and each sequential step in product execution briefly outlined.
    c. The teacher should give thorough and visually clear directions preceding each succeeding step of the art process.
    d. Individual daily growth should be clearly visible to the student. He should be able to see this growth taking place in the object.
    e. Students should be free to move within the work area.
    f. Only simple and obviously easy choices should be offered the student. As a result, creative activity must be held to a reasonable minimum and must be kept on the level of elementary problem-solving.
    g. Students should be given advance warning before being asked a question.
    h. Tool storage should be on an individual basis and should be fixed and unchanging.
7. His thought process appears to depend upon physical reality for its stimulation. As a result, he appears relatively unable to engage in abstract verbalization or abstract art production. Thus, additional art activities, involving subtle two-dimensional experiences in painting or designing should be kept to a minimum. Instead, his thinking should be stimulated primarily through physical experiences of a concrete and sensory nature. Verbalization of this kind of thinking should not be overstressed. It is preferable to look at his visual products as though they were the "words" through which he is attempting to communicate. The above "apparent ramifications" have been merely inferred from a study of the listing of 41 characteristics of the popular art activities of mentally handicapped children. Each point should be looked upon as a hypothesis which ought to be thoroughly tested before complete acceptance. In the second phase of this study, we hope to test out the validity of these assumptions.

Our "next steps" to be worked on during this coming year 1958-1959 will include:

G. Testing the validity of our "yardstick" to predict the popularity of new activities.

H. Testing the above seven "apparent ramifications."

University of Maryland Pursues Initial Study

On June 1, 1960, the University of Maryland received from the U.S. Office of Education a grant of $22,400 to finance the study described in the foregoing article. The 16 month research project expands the design to include the range of art for mentally handicapped pupils from primary through secondary levels. Use of the repertory grid has been continued to establish the characteristics present in popular activities at the younger age levels. The original hypotheses are being tested at the primary, elementary, and secondary levels. An additional task of pinpointing the most appropriate level for the use of each experimental activity now becomes evident.

Data will be provided by using an adapted curriculum for art instruction in twelve Special Education classes of the Arlington, Virginia Public Schools. Results of the adapted program in the six experimental classes will be compared with results of the traditional art instruction in the six control classes. The following measures will provide a basis for the comparison and a means of testing the hypotheses of the study:

1. Anecdotal records of behavioral reactions
2. Independent judgment of art products
3. Preference questionnaires
4. Grades in all subjects
5. Mental maturity test results
6. Achievement test scores
7. Teacher evaluation of student growth, behavior, and interest

Dr. Jean Hebeler, Coordinator of Special Education at the University of Maryland is director of the cooperative research project with Dr. Richard G. Wiggin supervising the art activities within the Arlington County special education classes.

FOOTNOTES

1. First phase study of mentally handicapped children enrolled in special junior and senior high school art classes in Arlington County, Virginia. Group age range: 13-17 years; Group median I.Q.: 75. Study conducted by art teachers Edna Boulware, Betty Blake, Roy Anderson, under the supervision of Richard G. Wiggin, Supervisor of Art, Arlington County Public schools, Arlington, Virginia.
2. All of the data gathering was done through the use of Dr. George A. Kelly's "Repertory Test Grid," reported in his book, The Psychology of Personal Constructs, Vol. 1, New York, W. W. Norton and Company, 1955, pp. 219-266.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

A STUDY OF THE SOCIO-ECONOMIC BACKGROUNDS OF PARENTS OF MIAMI UNIVERSITY (OHIO) STUDENTS

JOSEPH C. SEIBERT
Miami University
Oxford, Ohio
C. NEALE BOGNER
Miami University
Oxford, Ohio

THERE ARE many unanswered questions about the family backgrounds of college students currently enrolled in America's institutions of higher learning. Administrative officers and trustees responsible for the setting of fees and tuition frequently do so in terms of what other colleges and universities are doing. Too little research has been employed in evaluating the ability of students and parents to pay higher fees. Rather, there seems to be the assumption that there will always be a supply of candidates for higher education since dependence may be placed on the ever-increasing number of 18 to 21 year olds. This assumption has some validity because of the increased birth rate of the forty's. However, the matter of raising the cost of education may tend to create a college population of the financially elite. This group may, or may not, number enough to crowd or even fill existing, plus planned, facilities.

The colleges and universities which are assisted financially by the tax funds of municipalities or states need to be concerned about pricing education beyond the ability of students who qualify intellectually, if not financially. Trustees and administrative officials of these institutions must know the ability of students to pay the current and proposed cost of education. If this ability is less than suspected, there must be increased tax support from the legislative bodies responsible for the operation of these colleges and universities. With this in mind, John D. Millett, President of Miami University, proposed that a study be made of the social and economic backgrounds of students currently enrolled at Miami University in Oxford, Ohio.

A questionnaire, designed to obtain pertinent information, was mailed to every tenth student listed in the Miami University Student Directory for the academic year 1958-59. Of 571 questionnaires distributed, 555, or 97.2%, were returned. Students attending classes in the off-campus centers of the University which are located in near-by cities were not sampled. The sampling reflected the financial status of students living in University residence halls, fraternity houses, private homes, or in the village community, as well as a smaller number from near-by communities.

The adequacy of the sample can be demonstrated by a close correlation with actual percentages of student classifications known to exist in the total student population. In the sample 56.6% were men and 43.4% were women. Actual enrollment statistics indicate men as being 56.2% of the total enrollment and women as 43.8%. The correspondence in distribution by classes was likewise significant. Actual enrollment ratios indicated 33.1% as freshmen, 23.5% as sophomores, 19.2% as juniors, and 15.5% as seniors. The sample was composed of 34.6% freshmen, 23.8% sophomores, 24.2% juniors, and 15.3% seniors. Total enrollment statistics indicated that 73.8% were students from Ohio and 24.5% were from out-of-state families. In the sample, the findings indicated 73.0% as Ohio students and 27.0% as out-of-state students. In some of these proportions it may be noted that totals do not equal 100%. This is due to a few graduate and special students who were not included for the purpose of the ratios involved. The comparisons as given, nevertheless, are sufficient to indicate a degree of correspondence to permit the findings to be taken as adequate within the reliability of the sample size.

Tabulations of the responses received reveal important information for the benefit of University administrative decisions. For example, it was determined that about 25% of Miami students are gainfully employed during the academic year. The number of women students who work is relatively less than is true of all students and the ratio of out-of-state students who work is slightly below the average. Most of those who do work earn from less than $100 to not more than $400. In addition, the findings showed a substantial number of students earning board by working in the dining halls of the University or in fraternity houses. There was no effort to assign a monetary value to the amount of board received as compensation.

It is interesting to note, furthermore, that about three out of four Miami students work during the summer. There is less difference in this connection between men and women or between in-state and out-of-state students. However, men tend to work somewhat more than the average, as is likewise true of out-of-state students. Earnings during summer vacation periods range from $100 to less than $800 for most Miami students. About two-thirds earn less than $500. About four of each ten students plan to work during Christmas or Spring vacations. Of those who contemplated employment, most expected to earn between $50 and $200. Two-thirds expected to earn less than $100. These findings are summarized in Table I with a breakdown showing the different ratios and earnings figures for male students as contrasted with female students.

Table I illustrates in summary fashion that the greatest earnings during the school year represent the wages of male students. Average earnings for male students were $562 for the school year in contrast to $150 for female students. Likewise, during the summer period, men earned more than women but the differential is not as great. The average for all students during the summer period is lower than the average for all students during the school year because there are more women working during the summer and earning a relatively lower wage. There is some bias reflected in these averages which is caused by relatively high incomes of those regularly employed married students which influence the results both during the regular school year and during the summer and other vacation periods.

About one of every five Miami students holds a scholarship. More men have scholarship assistance than women and more Ohio students hold scholarships than is true of out-of-state students. About half of the scholarships have a value of $2,000 or less, figured over a four year period. Fewer than ten per cent of the students borrowed money to pay expenses during the academic year for which the survey was made. More loans were made to men than to women; three-fourths of the loans were for $500 or less.

Nearly 25% of the fathers of Miami students are owners or managers of small businesses. About one in eight are senior or junior executives in large businesses, and one in six are professional people. About one in eight are semiskilled or unskilled workers. The highest proportion of professional people and executives of large businesses is among the fathers of out-of-state students.

These findings with respect to occupations of fathers of Miami students have been gathered in Table II with the breakdown again shown for the fathers of male students as compared with the occupational classifications of the fathers of female students.

There seems to be little difference according to Table II in the occupational backgrounds of the fathers of the male students as contrasted with the fathers of the female students. To some extent, the male students of the University tend to come from families where the father is either a salesman or an administrative official in a large business. In contrast, the female students at the University are more likely to be from families where the father is an owner of a small business enterprise or where the father is in the classification of professional people. More male students are from families of skilled, unskilled and clerical workers; in contrast, more female students come from families where the father is retired, unemployed or deceased.

About one-third of Miami students have mothers who are employed. The ratio is somewhat greater for female students than for male students and more for resident students than for out-of-state students. One-fourth of the out-of-state students have mothers who work.

Slightly more than 75% of the fathers of Miami students have a high school education or better. About one-third have a college education. For out-of-state fathers, this latter ratio is 45%.

Table III reveals the educational backgrounds of the fathers of Miami students. From the table it is evident that girls come from better educated families than is true of the men at Miami. The table indicated that 82.9% of the fathers of the female students have a high school education or better whereas the same ratio for the fathers of the male students is 75.6%.

Occupational plans for most Miami students include other businesses or teaching, with about 20% of the students undecided.

More than 50% of Miami students selected the University, in part at least, because of its campus. Two-thirds checked academic reputation as a factor. The low cost of attending Miami was important to 50% and relatively more important to Ohio students than to out-of-state students, and to commuters more so than to students living on the campus.

According to Table IV, the cost of attending the University is more important to female students than to male student . The same is true with respect to the drawing power of the campus and the academic program. As would be expected, athletic programs of the University have much more appeal to the male students than to the female students. Word of mouth advertising by friends, associates and acquaintances is a more important reason for bringing the male students to the campus than is true of the female students. No doubt some bias entered in the checking of the academic program as a reason for selecting Miami because that is what one should be expected to consider as important. The bias is fairly consistent, however, and therefore the results are still indicative as to which segments of the student body consider the academic program more important than other segments.

Following the checking of six listed reasons for attending Miami, space was left for any voluntary response as to reasons for selecting the University.

TABLE I

EXTENT OF EMPLOYMENT AND AVERAGE EARNINGS OF MIAMI STUDENTS FOR SPECIFIED PERIODS DURING YEAR 1958-1959

Period	Male		Female		Total	
	Percent working	Average earnings	Percent working	Average earnings	Percent working	Average earnings
During Academic Year	32.8%	$562	18.8%	$150	26.8%	$499
During Summer Vacation	77.1	572	62.1	330	70.6	480
During Christmas or Spring Vacation	48.7	103	24.1	64	37.9	92

TABLE II

OCCUPATIONS OF FATHERS OF MIAMI STUDENTS

Occupation Classification	Male	Female	Total
Professional	14.4%	16.7%	15.4%
Officials, large businesses	14.4	12.5	13.5
Owners, managers, small businesses	22.0	25.4	23.5
Salesmen	7.0	4.6	6.0
Clerical	8.0	5.8	7.1
Skilled workers	11.2	9.6	10.5
Semi-skilled workers	9.2	9.1	9.2
Unskilled workers	4.2	2.5	3.4
Farmers	1.9	3.3	2.5
Military	1.3	1.7	1.5
Retired	1.6	2.1	1.8
Unemployed	.0	.4	.2
Deceased	4.8	6.3	5.4
	100.00	100.00	100.00

TABLE III

EDUCATION OF FATHERS OF MIAMI STUDENTS

Highest Level Reached	Male	Female	Total
Less than Grammar School	1.6%	.4%	1.1%
Grammar School	22.8	16.7	20.1
High School	41.5	43.3	42.3
College	34.1	39.6	36.5
	100.0	100.0	100.0

TABLE IV

RELATIVE IMPORTANCE OF SIX POSSIBLE REASONS FOR ATTENDING MIAMI UNIVERSITY

Reason	Male	Female	Total
Academic Program	64.5%	69.2%	66.5%
Miami Campus	51.8	60.8	55.7
Cost of Attending Miami	48.9	53.3	50.8
Friends' Suggestion	43.4	37.9	41.0
Parents Attended Miami	5.4	7.9	6.5
Athletic Program at Miami	10.2	.4	6.0

TABLE V

SOCIO-ECONOMIC LEVELS OF PARENTS OF MIAMI STUDENTS

Socio-economic Level	Male	Female	Total
Upper Upper	.0%	.0%	.0%
Lower Upper	.0	.4	.2
Upper Middle	29.0	28.2	28.6
Lower Middle	53.2	54.4	53.7
Upper Lower	17.8	17.0	17.5
Lower Lower	.0	.0	.0
	100.0	100.0	100.0

There was a number of reasons given ranging from certain academic programs to social activities and regulations. Most important, however, of the voluntary responses had to do with the location of the University. One in eight selected Miami because of its location close to home. This reason for selecting Miami was more important to female students than to male students.

Almost all students indicated the principal breadwinner in their families earned more than $4,000. Among Ohio students, between one-fourth and one-third had family breadwinners in the above $10,000 class, but this ratio was above 50% for the out-of-state students. Almost a third of the breadwinners supporting Ohio students earned less than $6000. Average incomes in families of female students appear to be somewhat lower than for male students. The findings for female students, interestingly enough, are less accurate since many more women students did not know about the incomes at home.

The questionnaire revealed that most Miami students and their families are in the lower-middle or upper-middle socio-economic groups. The sociologist would classify these families as being in the upper ranges to slightly above the ''common man'' designation. Almost three out of ten students are from fairly well-to-do families---families with members likely to be civic leaders and having civic responsibilities. About half of Miami students are from families just below this social designation where the parents are followers rather than leaders.

A substantial number of Miami students, almost one in every five, are from families which have only a minimum of social responsiveness as well as a minimum of discretionary income.

Table V summarizes these findings and reveals, at the same time, a striking comparison in the proportions of male and female students having similar socio-economic backgrounds. Variance in socio-economic backgrounds of Miami students are, therefore, more important in classifications related to resident students versus commuters and Ohio students versus out-of-state students than is true of female students versus male students.

As is true in almost every survey, questions raised by the tabulated results tend to provoke an awareness for further investigation and inquiry. The findings in this particular case are typical. Nevertheless, the project has gathered sufficient evidence to offer a substantial guidance in administrative decisions with respect to registration fees, and other financial requirements made of the students. In similar fashion, it is evident that the objectives of the University might be sharpened in view of a better definition of the characteristics of the student population being served.

A complete analysis of the questionnaire may be obtained by writing to C. Neale Bogner, Assistant to the President, Miami University, Oxford, Ohio.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

AN EXPERIMENT WITH THE CONTROLLED READER

FORBES BOTTOMLY
Spokane Public Schools
Spokane, Washington

A FEW years ago it was difficult to entice a reading expert into speaking a friendly word for any of the electronic devices which purported to teach reading skills. Even today there are those who remain unconvinced that the devices are more than tongue-twisting gadgets with catchy sales appeal for unsuspecting teachers and gimmick-addicted superintendents. The Traxler report of 1943 had left a lingering feeling that "controlled reading" and "metron-o-scope" aids were neither valuable nor valueless in reading.[1] Nevertheless, a growing number of modern reading specialists have discovered that certain electroinc aids, when carefully used, make marked contributions to developmental reading programs. Among such specialists are Betty Kiehn and Evelyn Moos of the Spokane Public Schools. For several years the two have been systematically testing various devices in actual classroom situations. In the Spring of 1959 they released the results of an experiment in which they used the Controlled Reader in the Spokane elementary schools.

The Controlled Reader is an electronic device which projects words, phrases, sentences and stories on a screen at controlled speeds. Its object is to force learners to acquire oculomoter habits which are associated with reading speed.

The Problem

The purpose of the experiment was to find out if a special short-term reading program involving the Controlled Reader might be more effective in boosting reading speed and comprehension than the regular developmental reading program of the Spokane schools. The investigators wished to know whether the device might have some value as a future integral part of the long-term developmental program. They set their experiment up to answer four specific questions. (1) On the average, will pupils gain more in reading speed under the Controlled Reader program than under the developmental program? (2) Will they gain more in reading comprehension, in vocabulary development? (3) On the average, will pupils who acquired skills under the Controlled Reader program retain them as well as those who acquired them under the developmental program? (4) Is scholastic ability any more a function of gain in reading ability under the Controlled-Reader program than under the regular program?

The Procedure

The experiment involved 460 pupils from two widely-separated Spokane elementary schools (see Table I). One of the schools, which we may call "School A", drew pupils almost entirely from an upper-middle socioeconomic neighborhood; the other, "School B" drew exclusively from an upper-lower socioeconomic neighborhood. In each school experimental and control groups were set up in the fifth and eighth grades. Experimental pupils included those with high and low as well as average scholastic abilities. Control pupils were matched with experimental pupils on the bases of age in months, scores on the Otis Quick Scoring Mental Abilities Test, battery scores on the Stanford Achievement Test, and sex. The null hypothesis regarding mean differences in age and test scores between the two groups was found acceptable at the .01 level of confidence in each school on each level.

The eighth-grade experiment began on October 13, 1958, and terminated on December 5. The fifth-grade experiment began on November 17, ending on January 9, 1959. Before the experiment began, immediately upon its termination, and five months after its termination different forms of the same reading tests were given to all participants. The Cooperative Reading Test was used with the eighth grade and the Chicago Reading Test with the fifth. The daily program for both groups was a 40-minute lesson which was conducted four times a week for the eighth grade and five times for the fifth. Similar programs and procedures were followed for all experimental and all control classes on each grade level. Close

---

[1] A. E. Traxler, "Value of Controlled Reading: Summary of Opinion and Research," Journal of Experimental Education, Volume XI, 1943, pp. 280-92.

TABLE I

NUMBER OF PUPILS PARTICIPATING IN THE READING EXPERIMENT

School	Grade 5		Grade 8	
	Experimental	Control	Experimental	Control
A	84	84	49	49
B	66	66	31	31
Total	150	150	80	80

TABLE II

REJECTION OR ACCEPTANCE OF NULL HYPOTHESIS AT .05 READING LEVEL REGARDING MEAN DIFFERENCES IN READING TEST SCORES BETWEEN EIGHTH-GRADE EXPERIMENTAL AND CONTROL GROUPS

Reading Test Category	School A		School B	
	Before Test	After Test	Before Test	After Test
Speed	Accept	Accept	Accept	Accept
Comprehension	Accept	Accept	Accept	Accept
Vocabulary	Accept	Reject	Accept	Accept

TABLE III

ACCEPTANCE OR REJECTION OF NULL HYPOTHESIS AT .05 LEVEL REGARDING MEAN DIFFERENCES IN READING TEST SCORES BETWEEN FIFTH-GRADE EXPERIMENTAL AND CONTROL GROUPS

Reading Test Category	School A		School B	
	Before Test	After Test	Before Test	After Test
Speed	Accept	Accept	Accept	Accept
Comprehension	Accept	Accept	Accept	Accept
Vocabulary	Accept	Accept	Accept	Reject

TABLE IV

ACCEPTANCE OR REJECTION OF NULL HYPOTHESIS AT .05 LEVEL REGARDING MEAN DIFFERENCES IN GROWTH FACTORS BETWEEN EXPERIMENTAL AND CONTROL GROUPS*

Reading Test Category	School A		School B	
	Grade 5	Grade 8	Grade 5	Grade 8
Speed	Reject	Reject	Reject	Reject
Comprehension	Reject	Reject	Reject	Accept
Vocabulary	Reject	Reject	Reject	Reject

* NOTE: Growth factor = amount of gain or loss between end of experiment and 5 months later.

supervision was maintained over each class by the investigators.

Experimental pupils, or those participating in the Controlled-Reader program, were divided into three groups within each classroom on the basis of their initial reading speed scores. Each group was rotated daily through three major phases of the program -- namely, preparation, reading practice, and testing. The preparation phase had two parts. Motivation which established the purpose and background for reading the selection was followed by Vocabulary Development which covered specific word meanings, multiple usages, recognition, synonyms, derivations, phonetic approaches, and dictionary use.

The reading practice phase entailed reading selections projected by the Controlled Reader. First, the story which had been read the preceding day was projected, then a new story was presented. After each story, the third phase, testing, took place via a comprehension check. Each daily lesson also included a short period of reading from a book at a level deemed comfortable for the pupils in terms of vocabulary.

In introducing the Controlled Reader to the pupils for the first time, the reading rate was intentionally set below the rate planned for them. It was slowly increased until a few pupils complained of its speed, then set back to a pace all could easily read. As the program progressed, the machine's rate of projection was increased whenever most of the pupils in a group reached or exceeded a goal of 80 per cent on the comprehension check. Those failing to achieve the goal were moved to the next lower group while those indicating a lack of challenge were moved to the next faster group.

Pupils in the control group proceeded with the regular developmental reading program. Developed under the leadership of Elementary Director, Marjorie Pratt, this is a long-term program to develop all aspects of reading ability, including values. Although the program has clear-cut, preplanned guidelines, there is no routinized daily ritual to which all teachers adhere. Instead, each teacher bases his approach largely on pupil interests. These, of course, are guided interests inasmuch as they are stimulated by leading ideas.

As interests intensify, children are led toward fundamental reading goals. They learn to follow directions, to recognize values placed on words and phrases. They are urged to search for "glittering" words and to identify dull ones.

Recognition of central thoughts and themes is emphasized as well as how to isolate important details. Locating of facts, finding proof, organizing data, generalizing, interpreting meanings, imagining, predicting outcomes and drawing conclusions are all part of the developmental reading process. Learning to skim, to use indexes, to develop precision and avoid carelessness are important. Skill in oral reading and intelligent listening are also goals of the program.

Developmental teaching procedures include the use of numerous regular and reference reading books as well as periodicals. Simple research ideas are formulated from pupil interests and projects are developed at appropriate levels of difficulty.

The Findings

On the reading tests given before the experiment began it was impossible to reject the null hypothesis at the .05 level regarding mean differences between the experimental and control groups in reading speed, reading comprehension and vocabulary. On the tests given immediately after completion of the experiment, it was found that although both groups gained appreciably in the three test categories, no significant mean difference between them was found at School A except in the vocabulary category in the eighth grade (See Table II). At School B, in the lower socioeconomic area, both the mean reading speed score and the mean vocabulary score of the fifth-grade experimental pupils was significantly higher on the test given immediately after termination of the experiment than those of the control group. (See Table III).

Of special interest is the finding that on both grade levels and in both schools the experimental students who scored in the upper third of their group on the Stanford Achievement battery displayed a significantly higher mean score in reading speed after completing the program than did control group pupils who also scored in the upper third. This was not evident in the scores of the before test. However, when the individual reading scores of the handful of fifth-grade pupils who read very rapidly (above 300 words per minute) were examined, it was discovered that all of these either scored lower, or nearly the same, on the after test than on the before test.

The most revealing data are found in what the investigators have termed "growth factors". Growth factors are gains in scores between the tests given immediately upon termination of the experiment and the alternate forms of the tests given five months later. Growth factors were used to judge retention of acquired skills. It was found with only one exception that the mean growth factors of the experimental group were greater than those of the control group and that the differences were significant (See Table IV). This would seem to indicate that the skills acquired during the Controlled Reader program were late blooming, revealing themselves after the pupils had returned to participate in their regular developmental reading classes.

Finally, to determine whether scholastic aptitude might be a greater function of gain in reading speed under the Controlled-Reader program than under the developmental program an analysis of variance was made. No significant difference was found between the two groups when those scoring in the upper third, middle third and lower third of their classes on the Otis Mental Abilities test were compared by this method. It was concluded that scholastic ability as measured was not any more a function of gain in one program than the other.

Summary

Researchers Kiehn and Moos concluded that the Controlled-Reader had demonstrated real promise for inclusion in the regular developmental reading program. They listed its major points of potential for the Spokane Schools as follows:

1. The major use of the Controlled-Reader should be in boosting reading speed, although it appears to have a beneficial, if delayed, effect on reading comprehension and vocabulary development.
2. Its best use is with average or better achievers who do not at first read rapidly. It should not, perhaps, be used with pupils who already read rapidly.
3. Since the experiment showed the greatest benefits of the Controlled-Reader program accruing after the program had terminated, it should be considered as a natural element of the long-term developmental reading program.
4. The Controlled-Reader program appears to have a more immediate effect upon reading speed in the fifth grade than in the eighth. It also appears to have a greater value in boosting reading speed among the pupils in the lower socioeconomic area of the city than among those in the upper socioeconomic area.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

A BEGINNING READING VOCABULARY

DANIEL W. FULLMER
Oregon State System of Higher Education
CLIFFORD J. KOLSON
Los Angeles State College

PERHAPS BECAUSE of its relationship to educational activities, the vocabulary of children has been of great interest to the educator. The size of preschool children's vocabulary was studied as early as 1907 by Doran (1). The Whipples (2), Mateer (3), Pelsma (4), Boyd (5), and Brandenburg (6), followed Doran with studies of the vocabulary of children.

One of the better known of the early studies of vocabulary was made by Thorndike (7). From an analysis of the reading vocabulary contained in children's literature Thorndike made a list of 10,000 words. A subsequent study by Thorndike (8) produced a word list of 20,000 words. In 1944 Thorndike with Lorge (9) produced a word list of 30,000 words.

One of the most significant studies was conducted by the Child Study Committee of the International Kindergarten Union (10). A word list of 2,596 words was compiled by making a frequency list from 893,266 running words spoken by kindergarten children and recorded by members of the committee. Generally, the early vocabulary of children studies were of a particular child. The IKU study was an attempt to get the vocabulary of a particular group of children. This study gave some indication as to the conceptual background of kindergarten children as revealed by oral language.

Other investigators studied the vocabularies of elementary school reading matter. Daw (11) produced a reading vocabulary for the middle grades. Horn (12) developed a basic writing vocabulary. Gates (13) compiled a reading vocabulary for the primary grades. Dolch (14) made his 220 word basic sight vocabulary.

All these studies have made contributions to the knowledge of the language and language development of children. The worth of these studies is aptly expressed by Figurel (15):

"The vocabulary investigations mentioned in this introduction, and others like them, have more or less assisted teachers and writers to produce materials that are somewhat within the comprehension of average children living in average communities."

However, just as broad general objectives are without value unless supplemented by specific objectives, so the vocabularies must be further refined so as to be applicable to a narrow area. Primary vocabularies become long lists of words much too cumbersome for the teacher to handle efficiently. What is needed is a small vocabulary for a specific area. With this in mind, the study reported on the following pages was conducted.

Preliminary Statement

With the increased use of the experience approach to beginning reading, a definite need exists for a word list to guide the teacher in Developing a word recognition necessary for success in beginning reading. In order for any list to be of value, it should satisfy certain criteria:

1. Size. The list should be small enough to be handled efficiently by teachers and by pupils.
2. Frequency of use. The list should contain words with which the child is most likely to come in contact in reading matter.
3. Level of application. The list should be designed for a particular grade level or area within a grade level.
4. Manageability. The list should be so constructed as to eliminate the need for the teacher to spend an excessive time in extracting the words she needs.
5. Utility. The list should have numerous applications in reading programs.
6. Consistency. The list should be formulated by the use of research techniques established by previous investigators in the area of inquiry.
7. Currency. The list should be based upon current data in the area.
8. Inclusiveness. The list should be derived from a major proportion of sources so as to justify the generalizations.
9. Significant. The list should correlate with the findings of other investigators in the area of inquiry.

Method of Study

This list of words has been compiled by tabulating the frequency of occurrence of individual words in eleven basal reading series. A total of forty-five pre-primers, primers, and first readers published between 1954-1959 were included in the study. This represented all the published basic reading series between the dates listed above.

The total list was then divided into five lists as follows:

- List I - Frequency count of 10 or 11
- List II - Frequency count of 9
- List III - Frequency count of 8
- List IV - Frequency count of 7
- List V - Frequency count of 6 or less

List V was dropped from the final total since it was felt that a frequency count of six or less would not make the word inclusive enough to be retained.

The list was then correlated with the International Kindergarten Union List and the Dolch Basic Sight Vocabulary. The matter of counting the different forms of the same word was resolved by considering root words and their derivitives as different words. This was done on the basis that we were concerning ourselves with a sight vocabulary and any assumption that the child could recognize a derivitive implies word attack skills.

The list compiled by this method follows.

The Vocabulary

LIST I

Word	IKU	D
a	x	x
and	x	x
at	x	x
away	x	x
can	x	x
come	x	x
fast	x	
for	x	x
go	x	x
had	x	x
has	x	x
have	x	x
I	x	x
in	x	x
is	x	x
it	x	x
little	x	x
make	x	x
now	x	x
one	x	x
said	x	x
the	x	x
there	x	x
up	x	x
with	x	x
you	x	x
out	x	x
LIST II		
all	x	x
are	x	x
be	x	x
big	x	x
box	x	
came	x	x
did	x	x

Word	IKU	D
do	x	x
down	x	x
from	x	x
fun	x	
funny	x	x
get	x	x
give	x	x
he	x	x
help	x	x
her	x	x
here	x	x
him	x	x
his	x	x
jump	x	x
like	x	x
look	x	x
looked	x	
may	x	x
me	x	x
mother	x	
no	x	x
not	x	
ride	x	
saw	x	
see	x	x
she	x	x
stop	x	x
they	x	x
this	x	x
to	x	x
too	x	x
two	x	x
was	x	x
we	x	x
went	x	x

Word	IKU	D
yes	x	x
LIST III		
am	x	x
as	x	x
asked	x	
baby	x	
back	x	
ball	x	
basket	x	
birthday	x	
black	x	x
blue	x	x
but	x	x
call	x	x
children	x	
could	x	x
day	x	
dinner	x	
dog	x	
eat	x	x
father	x	
feet	x	
find	x	x
girl	x	
going	x	x
goodby	x	
green	x	x
happy	x	
home	x	
house	x	
how	x	x
hurry	x	
into	x	x
kitten	x	
know	x	x
laugh	x	x
let	x	x
man	x	
may	x	x
new	x	x
of	x	x
on	x	x
put	x	x
ran	x	x
red	x	x
run	x	x
school	x	
something	x	
surprises	x	
take	x	x
that	x	x
want	x	x
wanted	x	
water	x	
where	x	x
who	x	x
will	x	x
work	x	x
yellow	x	x
your	x	x

Word	IKU	D
LIST IV		
about	x	x
after	x	x
another	x	
ask	x	x
book	x	
bump	x	
cake	xx	
called	x	
car	x	
cat	x	
daddy	x	
faster	x	
flower	xx	
found	x	xx
goat	x	
got	x	x
guess	x	
head	x	
high	x	
just	x	x
laughed	x	
let's	x	
liked	x	
made	x	x
many	x	x
milk	x	
must	x	
name	x	
noise	x	
oh	x	
other	x	
our	x	x
pets	x	
picnic	x	
played	x	
please	x	x
pony	x	
ready	x	
say	x	x
sleep	x	x
so	x	x
some	x	x
soon	x	x
street	x	
thank	x	x
them	x	x
then	x	x
three	x	x
time	x	
tree	x	
uncle	x	
us	x	x
wagon	x	
were	x	x
white	x	x

Analysis

In the section entitled "Preliminary Statement", certain criteria were established. Using these criteria as a guide in evaluating, we find:

1. Size. The final list contains 184 words. The wide use of the Dolch Sight Word List which contains 220 words would seem to indicate a list of 184 would be sufficiently small to be handled efficiently by pupils and teachers.

2. Frequency of use. The acceptance of a frequency of 50% is implied in the acceptance of the Dolch list. List IV represents words with the lowest frequency count. This count is in the 50% frequency count bracket.

3. Level of application. This list is aimed at the beginning reading level of the first grade; hence satisfies the criterion.

4. Manageability. The subdivision of the list into four lists determined by frequency gives the teacher a ready source of words with varying frequency without the time-consuming job of extracting from a master list.

5. Utility. The possible applications of this list are:

a. Testing. The list can be used to check the child's sight vocabulary as a means for determining readiness for formal reading instruction, for determining deficiencies in sight vocabulary, for establishing the group to assign a particular child.

b. Teaching. As a guide in the development of experience charts. As a guide to authors in preparing reading material appropriate to the beginning reading level.

c. Preventitive instruction. Once deficiency in sight vocabulary is determined, the list can be used for individual corrective instruction.

6. Consistency. Word count techniques have been established by educational leaders such as Thorndike, Lorge, International Kindergarten Union, Dolch, and others. The sources are given in the first part of this paper.

7. Currency. The criteria of currency would seem to be satisfied by the fact that basal readers published between 1954 and 1959 were included.

8. Significant. All the words included in this list are listed in the IKU list and hence are in the child's speaking and understanding vocabulary. Of the words in the list, 64% are included in the Dolch list. Since the Dolch list does not contain nouns, and since derivatives were considered separate words, the 64% figure represents the minimum figure. These facts would seem tp indicate that the list fulfills the criterion of significant.

The Basal Series Included in This Study

1. Allyn and Bacon
2. American Book Company
3. Curriculum Foundation
4. Ginn and Company
5. D. C. Heath
6. Houghton, Mifflin and Company
7. Lyons and Carnahan
8. The Macmillan Company
9. Row, Peterson Company
10. Scott Foresman Company
11. World Book Company

REFERENCES

1. Doran, E. W., "A Study of Vocabulary," Pedagogical Seminary, XIV (1907), 401-438.
2. Whipple, G. M., and Whipple, W., "Vocabulary of a Three Year Old BoyWith Some Interpretive Comments," Pedagogical Seminary, XV,(1908), pp. 1-22.
3. Mateer, Florence, "The Vocabulary of a Four Year Old Boy," Pedagogical Seminary, XV (1908), pp. 63-74.
4. Pelsma, J. R., "A Child's Vocabulary and its Development," Pedagogical Seminary, XVII (1914), pp. 95-142.
5. Boyd, William, "Development of a Child's Vocabulary," Pedagogical Seminary, XXI (1914), pp. 95-142.
6. Brandenburg, George C., "The Language of a Three Year Old Child," Pedagogical Seminary, XXII (1915), pp. 89-120.
7. Thorndike, Edward L., The Teachers Word Book of Ten Thousand Words, (New York: Bureau of Publications, Teachers College, Columbia University, 1921).
8. Thorndike, Edward L., The Teachers Word Book of Twenty Thousand Words, (New York: Bureau of Publications, Teachers College, Columbia University, 1932).
9. Thorndike, Edward L. and Lorge, Irving, The Teachers Word Book of Thirty Thousand Words, (New York: Bureau of Publications, Teachers College, Columbia University, 1944).
10. The International Kindergarten Union, A Study of the Vocabulary of Children Before Entering First Grade, (Baltimore: 1928).
11. Daw, Seward E., A Controlled Vocabulary for the Middle Grades, Unpublished Doctor's Dissertation, Pittsburgh: University of Pittsburgh, 1948.
12. Horn, Ernest, A Basic Writing Vocabulary, (Iowa City: Bureau of Publications, University of Iowa, 1926).
13. Gates, Arthur I., A Reading Vocabulary for the Primary Grades, (New York: Bureau of Publications, Teachers College, Columbia University, 1926).
14. Dolch, Edward W., Manual for Remedial Reading, (Champaign, Illinois, The Garrard Press, 1945).

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

THE RELATIONSHIP BETWEEN SCIENCE PROFICIENCY AND TRAINING AMONG LIBERAL ARTS GRADUATES

JACOB TUCKMAN
Philadelphia Department of Public Health
WILLIAM ETKIN
College of the City of New York

IN LIBERAL ARTS colleges, the basic required work of each department must be addressed increasingly to students whose major will usually be in some other field; e.g., required English composition and literature must be directed to non-English majors; history to non-historians, biology, chemistry and physics to non-scientists, etc. To determine the value and best methods of teaching these traditional subjects to the pre-professional and the non-specialist has become, in actual classroom practice, a dominant concern.

At the College of the City of New York, four one-semester courses, called Science Sequence, were organized in 1945 and offered as an optional choice to B. A. students in place of the standard required of two one-year courses of the "regular" departmental type. Each of the sequence courses offers lecture, recitation, and laboratory hours similar to those given in the regular departmental courses and attempts to present aspects and topics of interest to the non-specialist. These courses are: 1) physics, 2) chemistry, 3) biology, and 4) astronomy and geology. Although each course was given by members of the department of the science it chiefly dealt with, the courses were planned and controlled by a committee drawn from all of the science departments.

The purpose of this study is to compare the proficiency in science, as measured by the Graduate Record Examination, of City College graduates who had completed the sequence and those who had completed regular departmental courses. Such data were gathered as part of a broader study (1) in appraising the aims and objectives, teaching techniques, student-faculty relations, and effectiveness of the two types of courses in meeting the needs of the non-specialist. It should be pointed out that the Graduate Record Examination measures the fund of information accumulated by the student in a particular science, but does not measure the outcome of other course objectives, e.g., understanding of scientific method, application to human problems, historical development of the subject, etc.

From a list of liberal arts graduates in the five-year period, 1949 to 1953, the Educational Testing Service made available scores on the Graduate Record Examination for all students who had taken it. The number of students included 24 who had elected the sequence and 123 who had elected the regular departmental courses. The comparison between Sequence and Regular students was made by the paired comparison method, i. e., each Sequence student was paired with that student, from the entire pool of the regular students, whose overall scholastic college average most closely corresponded to his own. For each of the comparisons between Sequence and Regular on the Graduate Record Examination, the two groups showed no significant differences with respect to overall college average. Therefore, any differences in achievement on the Graduate Record Examination between the two groups cannot be attributed to differences in scholastic ability, but must be due to other factors, presumably the difference in courses taken.

The mean scores in physics, in chemistry, and in biology on the Graduate Record Examination for the Sequence and for the regular students are given in Table I. As might be expected, the data showed a tendency for score on the Graduate Record Examination to be affected by the extent of the student's course preparation. Comparisons which differentiated among the regular students according to the particular science they had taken in college showed that, in general, students who had taken a regular (1 year) course had a higher score than those who had taken the Sequence (1 semester) course in that subject. Correspondingly, those who had the Sequence course in a subject had a higher score than those with no college course in that subject. Also, those who had a high school course in a subject had a higher score than those with no high school or college course in that subject. It is important to note that the amount of improvement in score that can be ascribed to more preparation was remarkably small, representing only about a ten percent increase in each of the comparisons mentioned above.

However, the comparison between Sequence stu-

TABLE I

MEAN SCORE IN PHYSICS, CHEMISTRY, AND BIOLOGY ON THE GRADUATE RECORD EXAMINATION FOR STUDENTS COMPLETING SEQUENCE COURSES AND FOR THOSE COMPLETING REGULAR DEPARTMENTAL COURSES, PAIRED FOR COLLEGE SCHOLASTIC AVERAGE

Preparation of Regular Students	Number of Pairs	Regular Mean	Sequence Mean
Physics			
Unselected with respect to college sciences	24	488	493
One year of college physics	22	498	490
No college physics	20	458	491
No physics in high school or college	15	448	473
Chemistry			
Unselected with respect to college sciences	24	459	474
One year of college chemistry	18	536*	469
No college chemistry	23	435*	476
No chemistry in high school or college	13	375*	460
Biology			
Unselected with respect to college sciences	24	536	518
One year of college biology	23	599*	519
No college biology	17	472	526
No biology in high school or college	12	452	513

* Significant at < .01 level of confidence.

dents and regular students unselected with respect to the two one-year science courses they completed in college showed no significant differences between the two groups in mean score for physics, chemistry, or biology. Within the limitations of the data, therefore, it seems reasonable to say that there was no overall difference between Sequence and regular students with respect to achievement in science as measured by the Graduate Record Examination.

The data suggest that the particular science courses taken by liberal arts students are of less importance in the educative process than curriculum makers commonly assume. It would seem that the factual content of a course (as measured by the Graduate Record Examination) can safely be relegated to a secondary position in curriculum planning and other course objectives given primary importance.

REFERENCE

1. Report by the Committee on the Self Study Project, The College of the City of New York, 1955, Parts I and II. Dean Samuel Middlebrook, Chairman.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 7, March 1961)

THE EFFECT OF TRAINING IN LISTENING FOR SPECIFIC PURPOSES

SUE E. TRIVETTE
East Tennessee State College, Johnson City

AS THE CHILD gains more skills in the process of his education, he is expected to acquire the ability to listen effectively for various purposes. These purposes—to remember the main idea, to remember details, to make inferences, to get word meanings, and to follow directions—are also purposes for which he is expected to read. Definite training in reading for these purposes is being given throughout the elementary grades in public schools. However, no definite training in listening for these purposes is being given as a part of the school curriculum.

Statement of the Problem

It was the purpose of this study to determine whether definite training in listening for the "main idea", "details", and to make "inferences" affects the child's ability to listen for these specific purposes. As a subsidiary study and to facilitate the achievement of the primary purpose of the present study, attempts were made 1) to ascertain by means of a questionnaire those children who were considered "poor listeners" by teachers and/or parents prior to the administration of the pre-test; 2) to determine the relationship between the child's I.Q. and scores on the Listening Comprehension Test; and 3) to determine the relationship between reading level and scores on the Listening Comprehension Test.

Scope and Content of the Field Covered

Six fifth-grade classes, including 147 students at the George Washington Elementary School in Kingsport, Tennessee, were studied in an effort to determine the effect of definite training in listening for specific purposes.

The group was probably representative of a cross-section in range of ability, since no effort was made to group the students homogeneously. A wide range was represented as to socio-economic background, although students from a low socio-economic background were in the majority.

The materials used in this study included 1) two questionnaires entitled "Listening Information Form: Information From Parents", and "Listening Information Form: School Information", by Willard Abraham; 2) Listening Comprehension Test for Grades Four, Five and Six, by Maurice S. Lewis, and 3) daily listening exercises.

Items tested on the Listening Comprehension Test for Grades Four, Five and Six were 1) skill in listening for the main idea; 2) skill in listening to remember details; 3) skill in listening to make inferences; 4) skill in listening to understand word meanings; and 5) skill in listening to follow directions.

Although five skills were tested, training in listening was given in only three skills because these three skills—"main idea", "details", and "inferences"—are more suitable for group instruction. Developing skill in getting word meanings and following directions requires much more individualized attention.

The first thirty exercises from the Gates-Peardon Practice Exercises in Silent Reading, Book VI, were used as listening exercises. Two major factors influenced the choice of sixth grade level for these exercises: 1) Book V had already been used in some classes as silent reading exercises; and 2) the time at which the study began was near the beginning of the second semester in the fifth grade classes. Types A, B, and D were used, providing exercises on listening for the main idea, for remembering details, and for making inferences.

Findings

The highest possible score on the Listening Comprehension Test was fifty points; the lowest possible score was zero. The results on Form A, given prior to the daily listening exercises, were as follows: Median score 24.93, mean score 24.66 and mode 26.5. The results on Form B, given after thirty daily listening exercises, were as follows: Median score 28.87, mean score 28.63, and mode 24.5

The median score increased 3.94 points on Form B. The mean score increased 3.97 points. The decrease in modes points out the fact that some student scores decreased on Form B.

The Pearson r product-moment correlation between scores on Listening Comprehension Test, Form A and Listening Comprehension Test, Form B was .62. This coefficient of correlation is significant, showing the positive effect of training in listening. The number of students among 147 whose scores showed improvement was 114, or 77.5 percent. The scores of six students showed no improvement. These students comprised 4.08 percent of the total. The scores of 27 students, or 18.4 percent, showed a loss from Form A to Form B.

An analysis of questionnaires returned by parents and teachers revealed the following data:

1. Questionnaire results from parenta were compared with questionnaire results from teachers. Seventeen questionnaires of 112 returned by parents indicated that the child was considered a "poor listener". Of these 17 "poor listeners" identified by parents, only four were considered "poor listeners" by teachers. Eight of the "poor listeners" identified by parents scored below the fiftieth percentile on Listening Comprehension Test, Form A.

2. Questionnaire results from teachers indicated 29 "poor listeners". Two of the four "poor listeners" agreed upon by parents and teachers did not score below the fiftieth percentile on Listening Comprehension Test, Form A. Twenty of the students identified by teachers as "poor listeners" scored below the fiftieth percentile on Listening Comprehension Test, Form A.

3. Forty-six students were identified by parents and/or teachers as "poor listeners". Seventeen of these students did not score below the fiftieth percentile on the Listening Comprehension Test, Form A. Thirty-four additional students not identified by parents and teachers did score below the fiftieth percentile on the Listening Comprehension Test, Form A.

An analysis of I.Q.'s of students scoring below the fiftieth percentile on Listening Comprehension Test, Form A, revealed a range in I.Q.'s extending from 66 to 120.

1. The distribution of I.Q.'s among the group of 59 students scoring below the fiftieth percentile revealed a mean of 94 and a median of 94.5.

2. The distribution of I.Q.'s among the 29 "poor listeners" identified by teachers revealed a mean of 88.7 and a median of 87.7.

3. The distribution of I.Q.'s among the 17 "poor listeners" identified by parents showed a mean of 101.47 and a median of 100.

Separate correlations run between scores on Listening Comprehension Test, Form A and I.Q,'s, and scores on Listening Comprehension Test, Form B and I.Q.'s revealed the following data:

1. A correlation coefficient of .55 between I.Q.'s and scores on Form A.

2. A correlation coefficient of .5 between I.Q.'s and scores on Form B.

A correlation between reading grade levels from the Stanford Achievement Test and scores on the Listening Comprehension Test, Form A, revealed a coefficient of .61.

Conclusions

From the literature on listening, it may be concluded that through the teacher's good example, through questions designed for self-analysis by students, and through structural aids to meanings and key ideas, teachers should be able to give adequate guidance in listening for specific purposes if the listener also assumes his share of the responsibility.

The following conclusions may be drawn from the present study:

1. Training in specific listening skills was effective for most students included in this study. The present writer can suggest no possible reason for the fact that 29 students received lower scores on Form B than those received on Form A.

2. Parent and teacher judgment in identifying "poor listeners", singly or in conjunction, did not seem to be reliable in identifying "poor listeners".

3. The .61 coefficient of correlation between students' reading grade levels as indicated by the Stanford Achievement Test and students' scores on the Listening Comprehension Test, Form A, was significant, indicating a parallel relationship between reading skills and listening skills.

4. When specific listening comprehension skills, such as "main idea", "details", and "inferences" are improved, other comprehension skills, such as getting "word meanings" and "directions" tend to improve.

RESEARCH NEWS

John Schmid, Editor
University of Arkansas
Fayetteville, Arkansas

Readers are invited to send items of research interest to John Schmid, Professor, College of Education, University of Arkansas, Fayetteville, Arkansas.

••

WILLIAM S. GRAY

Dr. William Scott Gray, internationally known educator in the field of reading, died of injuries suffered in a fall from a horse on September 8, 1960. He was seventy-five years of age. At the time of the accident, Professor Gray was on vacation from his position as Director of Research in Reading in the Department of Education at the University of Chicago. He was Professor Emeritus of Education.

TEST STANDARDS

A committee of the AERA on Test Standards has been appointed to revise Technical Recommendations for Achievement Tests. It is proposed that the National Council on Measurements Used in Education and the American Psychological Association will cooperate in the revision of this manual.

TEACHER EDUCATION COLLOQUIUM AT IOWA STATE TEACHERS COLLEGE

The annual colloquium on Teacher Education was held on November 13-14 at Iowa State Teachers College. About forty superintendents of schools in Iowa are invited each year to come to the campus to discuss matters of mutual concern in teacher education. One of the main topics for consideration in the 1960 colloquium was the types of research by college staff members which would be most useful to public school personnel. Four current college projects in research and curriculum development were reported: Modern Mathematics Curriculum for Elementary and Secondary Schools, Group Counseling Techniques with Parents of Elementary School Pupils, Team Teaching in the Sixth Grade, and a special summer program for gifted children of the fifth-sixth grade level.

CREATIVITY

Research on "Creativity" by Jacob W. Getezels and Phillip W. Jackson, Educational Psychologists in the Department of Education of the University of Chicago, is the subject of an article in the University of Chicago Reports (Vol. 2, No. 1).

PROJECT GUIDANCE AND MOTIVATION OF SUPERIOR AND TALENTED STUDENTS

Dr. Clayton L. Bennett has become the Associate Director of the Project on Guidance and Motivation of Superior and Talented Students which is being carried out by the North Central Association of Colleges and Secondary Schools. The project seeks to identify, guide, and motivate students who rank in the upper twenty-five percent of mental ability and/or achievement based on national norms, and who by other agreed upon criteria are held to have ability comparable to those identified by such objective criteria. The purpose of this project is to find, develop, and implement procedures and programs in secondary schools which will carefully identify, encourage, and provide increased guidance services for the students.

IN-SERVICE TRAINING

Dr. Ned Flanders of the College of Education, University of Minnesota, is conducting a study to determine if an in-service training course has any effect on the spontaneous behavior of a teacher in the classroom.

RESEARCH AT THE UNIVERSITY OF FLORIDA

Several studies are underway in the College of Education of the University of Florida. Among these are:

1. "A Comparison Between College Faculty and Beginning Education Student's Perceptions of the Professional Role of Teachers and the Characteristics of Good Teaching." This study is being carried under the direction of Drs. H. Lewis and I. Gordon. It is expected to be completed in July 1961.

2. "Non-Intellectual Predicator of Achievement" is being undertaken by Dr. William Spears. A scoring scale has been established for the Bender Visual Motor Gestalt Test which distinguishes between levels of academic achievement for white girls. Purposes of the study are to answer the questions of why the scale predicts for only one sex and race, and which academic areas can be predicted best.

3. "Learning as a Function of the Amount Learned." The study is being carried on by Drs. Webb and Spears. The purpose is to investigate the relation of the speed of learning to the amount of material learned.

4. "The Study of the Relationship of Children's Perceptions to Achievement and Behavior." This study is being carried on by Dr. Combs. It is attempting to study how children see themselves in the world in which they live as they progress through school.

CHILDREN'S SELF-CONCEPTS

Drs. Joseph C. Bledsoe and Carl C. Garrison of the College of Education of the University of Georgia have received a two-year grant from the U. S. Office of Education for studies in the self-concepts of elementary school children. The self-concepts will be studied in relation to their academic achievement, intelligence, interest, and manifest anxiety.

HANDWRITING RESEARCH

The Handwriting Foundation desires to contact investigators who are engaged in research in the area of handwriting, those who might become involved in such work, and those who may not want actually to work in this area but just want to keep informed about it. The Foundation is interested in obtaining this information not only from persons in colleges and universities but also from those in public schools, consultancy, industry, and government functions. Further information may be obtained by writing to Handwriting Foundation, 1426 G Street, NW, Washington 5, D.C.

"CREATIVITY" RESEARCH

Workers in the field of creativity research are invited to send their findings to the ETS Creativity Research Exchange. The Exchange, which has been in operation for two years, serves merely as a clearing house: it reproduces and circulates to members research data which might not become available in professional literature for some time. To obtain standard forms for outlining findings, address Dr. Samuel Messick, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

TV EVALUATION QUESTIONNAIRE

In a project supported by Fund for the Advancement of Education, a TV evaluation questionnaire, entitled "Sizing Up Your School Subjects", has been prepared by Drs. H. S. Dyer and A. H. Ferris to measure the reactions of high school students to instruction by the medium of television. It is so designed that it may be given in a way which prevents the student from being aware that his answers will be used to assess his reactions to TV instruction as such. It was administered in a large city high school system to biology students last year and this year in a group of large city schools. The results showed that conventional instruction was preferred over TV instruction.

A Dembar Publicatio

The Journal of Educational Research Box 1605, Madison 3, Wisconsi

Devoted to the scientific study of education

YOU'RE AN OLD FRIEND OF THE FAMILY,
SO YOU'RE THE FIRST TO KNOW:

Yes, as a reader of The JER, we want you to be the first to know that The Journal of Educational Research is about to add to its size and scope--with more pages and new features--the better to bring you rapidly and thoroughly the latest developments in basic investigations and advanced practices throughout education.

Beginning this fall The JER will present each month a special page on "What's New in Educational Research"--an up-to-the-minute digest for the reader who wants rapid orientation and insight. Plus more full-length papers by distinguished researchers, and other editorial departments we're confident you'll find both interesting and significant.

You can make A SUBSTANTIAL SAVING by entering or renewing your subscription now before our rates go up.

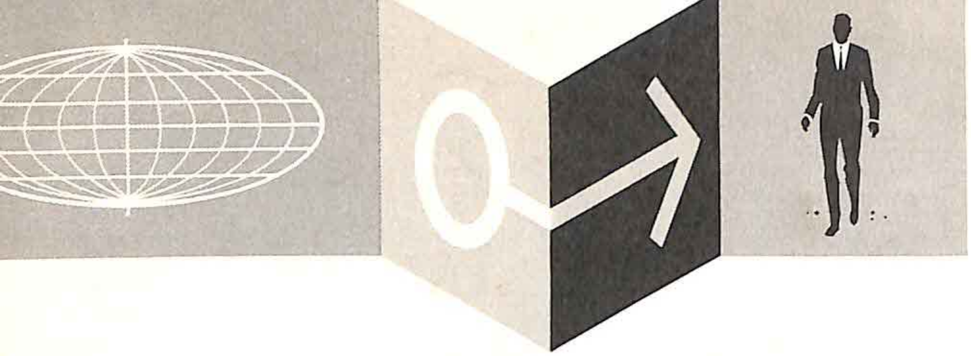
A handy reply card is attached, along with two pages describing The new JER in more detail.

Sincerely yours,

A. S. Barr

A. S. BARR
Editor

PS: If you receive another announcement of our special offer, we'd appreciate your passing it on to a colleague.



The World, Th

At the growing edge of productive scholarship

THE WORLD OF EDUCATIONAL RESEARCH IS EXPLODING

It is almost commonplace to say so, but it is true. The growth of educational research is almost incomprehensible in its size and impact. In the past 10 years, more time, talent, and money have been invested in pushing outward the frontiers of educational knowledge than ever before in the history of man. In those same 10 years the findings of productive scholarship have had a profound influence on the scope and character of educational programs and methods. In the next 10 years we can confidently expect even more significant developments.

Devoted to the scientific study of education

THE JOURNAL OF EDUCATIONAL RESEARCH IS GROWING, TOO

In scope, in size, and in stature The Journal of Educational Research has sought constantly to improve its services to its chosen field. Over the past 15 years we have pioneered new formats that have made possible the more rapid and economical reproduction of more learned papers than had hitherto been thought possible. Now we are about to take another marked step forward — with even more pages devoted to selected research reports, new editorial features of added appeal and utility, and a revised typography that typifies the forward-looking role of The JER.

In Investigation and Interpretation

FOR 40 YEARS THE "MOST-QUOTED" JOURNAL

The Journal of Educational Research is the *first* journal in its field — the oldest, the most frequently cited in professional literature. Since 1920 The JER has been the recorder of progress in education. In that period the "frequency of use" of the major educational publications has twice been studied. Both times The JER ranked No. 1. In the past five years JER circulation has almost doubled. We think these a striking recognition of JER acceptance and coverage.

The scholarly journal with the common touch

FROM BASIC RESEARCH TO ENLIGHTENED APPLICATION

We can assure you that The Journal of Educational Research is not just another journal. It bridges the gap between fundamental investigation, applied scholarship, and practical application. The Journal reports findings, yes, but it also illustrates the ways that alert educators can study their own problems and devise improved practices. For both the technical work and the field worker The JER is a mine of ideas.

In the universe of education

A RUNNING RECORD OF LEADERS AND LEADERSHIPS

Traditionally the "big names" in education have added substance to The JER.

Through the years The Journal of Educational Research has presented early evidence on all the major break-throughs in education: individual differences, learning and learning problems, training techniques, tests and measurements, curriculum, counseling and guidance, methods of teaching in all subjects at all levels, supervision, investigations of reading, teacher education, evaluating teacher effectiveness, administration, the winning of public support, and other areas.

Journal, and You

By and for educators at all levels

WIDE OUTLOOK FOR A WIDE READERSHIP

The Journal of Educational Research belongs to no one country, no one institution, no one association. Its outlook is not circumscribed by geography or slant. The JER lives at the growing edge of productive scholarship in the universe of education, devoted to the discovery, documentation, and dissemination of the knowledge and insights by which new truths are found.

The readership of The JER testifies to its broad significance. Leading libraries are represented almost without exception. So are college and university administrators, professors of education, psychologists, elementary and secondary school specialists and teachers, vocational and agricultural education personnel, adult education leaders, research workers, educational associations, armed forces educational advisors, students, interested laymen.

Nearly 20% of The JER circulation goes overseas.

Discovery, documentation, dissemination

TOMORROW'S EDUCATION TODAY

Perhaps you used to be able to keep abreast of educational developments by reading annual abstracts or perennial digests. But no longer. The field of education is too swift and complex for that. Today's experiments become next week's concepts and next month's practices. Only by making up-to-the-minute research your "must" reading can you hope to stay in the forefront of your profession.

In every field there is a journal that stands for excellence, whose regular reading is a requisite. Now more than ever before, in the expanding universe of educational knowledge, *the* journal is The Journal of Educational Research.

Features:

Nine issues around the calendar: September, October, November, December-January, February, March, April, May-June, July-August.

Twenty percent more pages devoted to the latest research papers.

"What's New in Educational Research" . . . a monthly digest of late developments for the reader who needs rapid orientation and insight (and who doesn't?).

Thought-provoking editorials on emerging trends, problems, break-throughs.

News Summary . . . a page of highlights and sidelights on conferences, grants, appointments, projects.

Book reviews and article citations . . . providing a running over-view of the entire field of Education.

Critiques . . . "think pieces" by leading educators on the state of the discipline.

Look for these and othe

VITAL REPORTS

COMING IN SEPTEMBER: *How to Evaluate Teacher Performance.* A remarkable synthesis of the latest literature on this critical subject.

COMING IN FEBRUARY: *Why Johnny Can Read.* A collection of current papers on a topic of perennial significance.

COMING IN MAY: A special issue devoted to *Teacher Personnel Problems.*

PLUS periodic reports of new research on testing, learning, curriculum, administration, guidance, identifying the exceptional student, and many other areas.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 APRIL, 1961 NUMBER 8

CONTENTS

Rationale of Norms on an Adult Personality Test, the 16 P. F.—For American Women—Raymond B. Cattell and Ronald R. Greene 285

Peer Status and Reading Achievement—O. V. Porterfield and Harry F. Schlichting 291

Devereux Readability Index—Edgar A. Smith 298

Relationships Between the Lee-Clark Reading Readiness Test and the 1947 Revision of the Stanford-Binet Intelligence Test, Form L—K. M. Parsley, Jr., and Marvin Powell 304

Oral Reading Used as an Indicator of Reactions to Frustration—Gladys Natchez 308

What Understandings Do Second Grade Children Have of Time Relationships?—J. D. McAulay 312

The Aliferis Music Achievement Test as a Predictor of Success in Music Theory at St. Olaf College—Adolph White 315

Research News—John Schmid 318

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

RATIONALE OF NORMS ON AN ADULT PERSONALITY TEST, THE 16 P.F. -- FOR AMERICAN WOMEN

RAYMOND B. CATTELL
University of Illinois
RONALD R. GREENE
Ohio Wesleyan University

ALTHOUGH THE validity of a test is more important than its standardization, since valuable work can be done in research without an accurate standardization, yet there comes a time when the more extended use of a well constructed test requires attention to the finer points of representative standardization.

The Sixteen Personality Factor Questionnaire, which is the only factored personality questionnaire based on fully attained oblique simple structure, and an array of behavior representative of the total personality sphere, was released as a research instrument just ten years ago. Now that its factor structure has been checked in two independent researches (2, 4), and the unitary nature of its factors has been witnessed by clinical and general evidence (5, 6, 7), there is increasing demand that its temporary standardization shall be replaced by one having as much care in its calculation as went into the original obtaining of the factor valid items themselves.

So far, the standardization may be considered adequate in regard to University students, but probably not adequate in regard to the general adult population. The present report describes the procedures in what we regard as a more adequate standardization, and brings out certain dangers in common procedures on which we and others have hitherto depended.

Up to the time of the present article, the "general adult" standardization (as distinct from the "undergraduate student" standardization) was derived from data obtained in the course of testing adults gathered in a variety of occupations, principally for the purpose of establishing, for vocational guidance and selection practice, the differences in personality profiles which exist among persons stably settled in various occupations.

The principle followed in deriving general population norms from data for occupational groups required that we 1) classify the occupations according to the ten major categories in the U.S. Census Report, with the help given by Shartle's work on classification of occupations (9), and 2) that we multiply the number of individuals in all but the largest category by the ratio necessary to bring the numbers in our samples from the main occupational groups to the same proportions as those existing in the country as a whole.

This principle of "stratified sampling according to main occupational categories" is admittedly a rough one. It does not insure that the age distribution correspond to that in the general population, nor can one be sure that certain narrowly specific occupations that one happens to be able to test are a random selection from the broader occupational category which they are allowed to represent. For example, we used airmen to represent the military group, though they may be somewhat differently selected from army and navy personnel.

However, the literal "man on the street" is the hardest possible subject to get and extremely few psychological measures are really based on him. A scrutiny of such surveys of mental tests as Buro's Mental Measurement's Year Book will show that all but a minority of tests for adults actually describe standardizations only for students, and that still fewer gave an adult standardization based on a proper sampling of the adult population in occupation, age, and region of the country. Although we believe that our previous principle of "stratified sampling according to occupational category" is one of the few available ways of getting at the general adult population, it admittedly still misses the unemployed and persons in uncommon occupations, and, as we found in the standardization of the Cattell intelligence test, twenty years ago (1), it turns out that the higher social status occupational categories are almost invariably more reliably sampled than the lower.

In the case of adult women, a majority of whom are in the home, the sampling by occupations, however, becomes inappropriate and open to a substantial systematic bias. Our results below will show the adult female population sampled in this way differs significantly from one sampled by better methods. This article is concerned with improved techniques and with resulting new 16 P. F. norms for adult women. In effect we continued to derive norms by the above occupational sampling method for a group of women largely representing single women and then proceeded to a distinct and specially worked out sample of women in the family and finally combined the two. For the data on women in families we are indebted to Mr. Howard Trumbull, president of National Family Opinion, Inc., which has used the 16 P. F. test in certain important studies, and has accumulated data on a carefully planned and unprecedentedly adequate stratified sampling of the adult female population of the United States. That the NFO sample, used in opinion and marketing research by various large business concerns, is an appropriate stratified sample of the total population of the United States with respect to a proper balance on age, region of residence, social status, etc., will be evident from Table I.

Since the NFO sample is reached by mail, the psychological question which immediately arises is "Will individuals who answer personality questionnaires by mail respond to them in the same way as in a supervised testing situation?" It will be noted that we have said "in the same way," not "as correctly as," for, as Cattell has pointed out in connection with intelligence tests (6), there are even advantages, in terms of reliability, to the answering of tests in an impersonal situation. Indeed, in a different questionnaire from the present one, it has been shown that the answers in a face-to-face situation are somewhat less reliable than those obtained when the questionnaire is done away from the examiner. However, in the present case it was decided to test this assumption first, and in order to do so, thirty-four subjects were taken in the Toledo area who first answered the questionnaire in a personal test situation and then answered the same form again after an interval of two weeks by mail.

No significant difference was found in score leads or reliability coefficients across methods compared with within methods. Of course, the NFO respondents are accustomed to answering, in responsible fashion, various questions sent to them by mail, so it may or may not be true that in general questionnaires are as reliably answered one way as the other, but for our present purposes, we are safe in proceeding with testing by correspondence.

With this experimental reassurance, NFO proceeded next to send out the IPAT 16 P. F. Test, Form A, to 1,000 respondents very carefully chosen to distribute according to the requirements of a true stratified sample of the U.S. female population, as shown in Table I. At the same time, when tabulations were finally made, 804 usable questionnaires had been returned, and the means and sigmas of resulting raw scores on each factor came out as shown in Table II.

A comparison of these means with those indicated in our tentative standardization (see Table 4a in mimeographed supplement to (8)), on 215 non-student women will show several significant differences. It is not a question of which norms are "right"—for some deviation from the ideal can be pointed out in either—but rather a question of what combination of the two samples will give the best estimate of the general population values. The main difference is that our earlier sample was obtained by the first described method of sampling occupations, and consists largely of single women in a full time calling, whereas the NFO sample is mainly married women at home. The NFO sample is a great improvement on our first sample in many ways, notably the control of region, age groups and income groups, but since only four fifths of the perfect sample responded, there is some selection by readiness to "volunteer", and one suspects a more subtle selection in that people who agree to work on such a panel without pay may have some personality "need to belong" or even to depend.

Regardless of possible influences—and it would constitute a second study with valuable indications for social psychology to explore these—the fact is that the differences which exist in the primary factors make psychological sense in their consistency of representation of second order factors. The mainly single, fully employed women are less neurotic and more masculine (8) than the housewives. This by-product of our inquiry should interest social psychologists concerned with the ecology of selection and situational influences, but we must return to our problem of combination.

Insofar as the present article is a discussion of principles as well as actual example of standardization, one can recognize that ideal procedure would be to combine our samples in such a way that the proper frequencies according to the census are restored. Incidentally, it is no criticism of the NFO sample that it is short of single women, for its purpose is to reflect family buying power, and for women-with-families the proper stratification of its sample has been carried very far indeed. A completely artless combination of the two samples cannot be made, for when samples are differently biased on different parameters, obviously no single ratio will bring all to the ideal, and, moreover, we suspect some biases to be important even though they cannot be quantified.

For example, there may be some selection of the NFO sample, as suggested above, toward more dependent personalities, and we suspect that because our occupations were contacted through students, they are biased in the direction of having college students temporarily in them, e.g., the waitresses were around a university town.

Accordingly, we cannot claim as yet a perfectly stratified sample, but only a closer approach thereto than before, and in this approach the most substantial thing we can do is to aim at proper proportions of married and single, housewife and occupationally attached groups. An examination of the Statistical Abstract of the U. S. Census, with extrapolation of the 1958 figures to 1959 for discernible trend gives the proportions for ages 20 to 65 ap-

TABLE I

COMPARISON OF PERCENTAGES IN THE STRATIFIED SAMPLE AND IN THE TOTAL UNITED STATES POPULATION

Geographic Divisions	% in U. S. Population	Number in NFO Sample of 1,000	% in NFO Sample
New England	5.7	57	5.7
Middle Atlantic	19.5	195	19.5
South Atlantic	14.2	142	14.2
East North Central	20.7	207	20.7
West North Central	9.0	90	9.0
West South Central	9.5	95	9.5
Mountain	3.7	37	3.7
Pacific	10.7	107	10.7
Population Densities			
Rural	24.0	240	24.0
Cities 2,500 - 499,999	15.4	154	15.4
Metropolitan Areas			
50,000 - 499,999 (Central cities)	10.2	102	10.2
50,000 - 499,999 (Outside cities)	8.7	87	8.7
500,000 - 1,999,999 (Central cities)	9.0	90	9.0
500,000 - 1,999,999 (Outside cities)	7.2	72	7.2
2,000,000 and over (Central cities)	13.7	137	13.7
2,000,000 and over (Outside cities)	11.8	118	11.8
Age Groups (Homemakers)			
Under 25 years	9.9	98	9.8
25 - 34 years	24.3	243	24.3
35 - 44 years	23.1	232	23.2
45 - 54 years	17.7	177	17.7
55 years and over	25.0	250	25.0
Family Income Groups			
Under $3,000	23.2	230	23.0
$3,000 - $3,999	10.7	107	10.7
$4,000 - $4,999	15.6	156	15.6
$5,000 - $5,999	13.6	136	13.6
$6,000 - $6,999	10.7	108	10.8
$7,000 - $9,999	17.4	174	17.4
$10,000 and over	8.8	89	8.9

TABLE II

MEANS AND SIGMAS OF 16 P. F. RAW SCORES ON THE STRATIFIED SAMPLE OF 804 WOMEN

Personality Factor	Mean Raw Score	Sigma of Raw Score
A	9.1	3.1
B	7.4	1.9
C	16.2	3.4
E	11.8	3.6
F	11.8	4.1
G	13.2	2.7
H	11.5	5.1
I	10.9	2.8
L	8.0	3.5
M	11.0	3.2
N	9.8	2.6
O	11.3	3.8
Q_1	7.9	2.8
Q_2	10.2	3.4
Q_3	10.2	2.9
Q_4	13.5	5.0

TABLE III

DISTRIBUTION (AGES 20 TO 64 YEARS) OF MARRIED VS. SINGLE, WORKING OUT VS. AT HOME OF U.S. WOMEN

	Married	Single (Widowed, Divorced)
	37.6 million (79.6%)	9.5 million (20.4%)
Working Outside Home 23.1 million (49.0%)	15.5 million (32.8%)	7.6 million (16.2%)
Working at Home 24.0 million (51.0%)	22.1 million (46.8%)	1.9 million (4.2%)
Total..		47.1 million

TABLE IV

MEANS AND STANDARD DEVIATIONS OF 16 P.F. RAW SCORES FOR BEST APPROXIMATION TO STRATIFIED SAMPLE OF U.S. WOMEN

Personality Factor	Mean: Raw Score	Sigma of Raw Score
A	8.96	3.11
B	7.70	1.70
C	16.62	3.34
E	12.10	3.30
F	12.72	4.33
G	13.38	2.81
H	12.00	5.05
I	10.60	2.92
L	7.68	3.50
M	10.82	3.23
N	9.92	2.60
O	10.56	3.80
Q_1	7.96	2.76
Q_2	9.96	3.45
Q_3	10.38	2.90
Q_4	12.76	5.10

proximately as shown in Table III.

Of the NFO sample, 92% are married and 27.8% are working out (16.7% full time, 11.1% part time) and of our occupational sample, obviously 100% are working out (roughly 75% full time) and approximately 80% are single. To have a correctly stratified sample of U.S. women in regard to being married or single, the NFO and occupational samples should be put together in the ratio of five to one, (actually 100 to 21), while to get the working out/working at home representative correct, we should add two-fifths of the occupational sample to one of the NFO.

Since the former is better ascertained (degrees of employment being hard to determine), we compromise on a ratio of 4:1 (NFO to occupational sample). To go to any finer figures would be unrealistic in terms of the size of samples and what is known about them; but this combination is a definite improvement on either sample and the best estimate obtainable from our samples of the ideal stratified sample of American women. Table IV shows the means and sigmas from combining the groups in these proportions.

These are the norms being supplied with the 1960 edition of the Handbook Supplement for the 16 P. F. Questionnaire, which will remain in force until results on a longer stratified sample can be obtained. A parallel report will be published for non-student men. Meanwhile it is interesting to note that this new standardization for women agrees with the first tentative standardization in showing differences of women from men significant beyond the P = .01 level in that women are cyclothyme (A factor), less dominant, (E factor), more premsic (protected emotional sensitivity; I factor), and more guilt-prone (O factor).

The writers wish to express their indebtedness to National Family Opinion, Inc. and the Institute of Personality and Ability Testing for material contributory to this article.

REFERENCES

1. Cattell, R. B. "Occupational Norms of Intelligence and the Standardization of an Adult Intelligence Test," British Journal of Psychology, 1934, 25, 1-28.
2. Cattell, R. B. "The Main Personality Factors in Questionnaire, Self Estimate Material," Journal of Social Psychology, 1950, 31, 3-38.
3. Cattell, R. B. "A Shortened "Basic English" Version (Form C) of the 16 P. F. Questionnaire," Journal of Social Psychology, 1956, 44, 257-278.
4. Cattell, R. B. "Validation and Intensification of the Sixteen Personality Factor Questionnaire," Journal of Clinical Psychology, 1956, 12, 205-214.
5. Cattell, R. B. "Second Order Personality Factors in the Questionnaire Realm," Journal of Consulting Psychology, 1956, 20, 411-418.
6. Cattell, R. B. Personality and Motivation Structure and Measurement, World Book Co., New York, 1957.
7. Cattell, R. B., Day, M. and Meeland, T. "Occupational Profiles on 16 Personality Factor Questionnaire," Occupational Psychology, 1956, 30, 10-19.
8. Cattell, R. B. and Stice, G. F. The Sixteen Personality Factor Questionnaire, Institute of Personality and Ability Testing, 1604 Coronado Drive, Champaign, Ill.
9. Shartle, C. L. Occupational Information: Its Development and Application, New York: Prentice Hall, 1946.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

PEER STATUS AND READING ACHIEVEMENT

O. V. PORTERFIELD*
San Fernando Valley State College

HARRY F. SCHLICHTING*
Ohio Wesleyan University

SOCIAL ADJUSTMENT of pupils is one of the important aims of education. In the last half century, increased attention has been given this factor by workers in the field of education, medicine, and social work. Good emotional adjustment, or the lack of it, has been found to influence a person's physical well-being, his educational progress, and his attitudes toward society as a whole. While these relationships in broad areas have been fairly well established, many relationships within this context have been left largely unexplored. This paper is concerned with one of these areas, viz., the relationship between status in reading achievement and status in peer prestige.

The Problem

The probability that some relationship exists between social adjustment and reading achievement is indicated by the somewhat similar personal characteristics which have been ascribed to persons with certain peer status and to persons with certain reading achievement status. For example, research studies have shown that pupils who have low peer prestige status also tend to possess the following personal characteristics: a low sense of personal worth and unusual anti-social behavior (9); talkativeness and restlessness (7); and their energies frequently are directed in such a manner as to interfere with or frustrate the activities of the group (13). Similarly, the characteristics of retarded readers have been described as: aggressive and cocky, significantly less insightful and less likely to accept or acknowledge blame, and frequently hostile toward and resentful of authority. (19).

This similarity in personal characteristics appears even more pronounced for those at the other extreme in peer prestige and reading achievement. Research has indicated that the pupils who have high peer prestige status also tend to possess the following personality traits: cheerfulness, friendliness, and enthusiasm (7); greater participation and striving for the approval of the group, sensitiveness to the feelings of others (15); and often are in leadership positions acceptable to society (4). Successful readers have been found to display a similar pattern of personality characteristics which have been described as: superior in persistence, self-control and concentration of attention (8); more introspective (18); and generally enjoy good rapport with the environment (19).

While these descriptions of personality characteristics might be variously interpreted, a broad interpretation of these and other studies formed the basis for the major hypothesis of this study, viz., that a significant relationship exists between peer prestige status and reading achievement status. To determine the effect of different variables upon this relationship, the following minor hypothesis was formulated: the relationship between peer prestige and reading achievement holds true irrespective of (a) the area of prestige, or (b) the socio-economic level of the school community.

The Procedure

To test the major and minor hypotheses necessitated the operational definition of terms and concepts, the selection of instruments to be used, and the identification of the population.

Definition of Terms

Peer prestige status was defined as the pupil's specific rank which resulted from his being chosen or rejected by the peers of his class. Since the specific area utilized to determine prestige status undoubtedly influences the results of a sociometric study, the bias of a single factor was avoided. Consequently, peer prestige status was interpreted to represent a composite score of prestige status in four rather well-defined areas in which leadership might be exercised, viz., social acceptance and rejection, physical agility and development in the form of sports leadership, sensitivity to and willingness to help meet the needs of a group or situation in the form of political leadership, and ability and desire to excel in school work in the form of academic leadership. Each of these constitutes one of the areas of prestige identified as part of the minor hypothesis and the sum total of these factors should indicate, with a considerable degree of reliability, the status accorded a person by his peers.

---

*O. V. Porterfield formerly was Elementary School Principal, Tulsa, Oklahoma, and is now at San Fernando Valley State College, Northridge, California. Harry F. Schlichting formerly was at the University of Tulsa and is now at Ohio Wesleyan University, Delaware, Ohio

Reading achievement status is defined here as the rank which a pupil attained on a standardized examination in reading, his status being fixed by his specific rank when all scores for each class were arranged in rank order. Thus a pupil who scored a high mark on a reading achievement test would have high status in reading achievement and a pupil who scored a low mark would have low reading achievement status.

Instruments Used

To ascertain the status of pupils in reading achievement and peer prestige, measuring instruments had to be selected which would provide accurate units which could be utilized to ascribe status to pupils. A standardized reading achievement test was selected to determine status in the area of reading but a new sociometric instrument had to be constructed which would provide the pupils with an opportunity to attain status in the selected areas as well as provide an index of peer prestige.

The Stanford Achievement Test, Intermediate Battery Partial, Form JM (6) was administered by the classroom teachers under the supervision of the building principal to all sixth grade pupils in the entire Tulsa school system during the period of March 4 to 15, 1957, as a regular annual procedure in the testing program of this school system. The reading achievement scores for this study were obtained from the results of Test I, Paragraph Meaning, of this Stanford Achievement Test. The actual performance scores for reading achievement were used in this investigation since a pupil's actual achievement would be the true index of his status within the group in the area of reading.

A new multiple-factor sociometric questionnaire was constructed to determine the peer prestige status of the pupils. This instrument was constructed in accordance with the criteria suggested by Northway (14) when she wrote: "The areas selected should include different aspects of possible association such as work, play, visiting; and some of them should assume a continuity of activity."

The sociometric questionnaire was constructed to include the different areas of prestige which were defined as components of peer prestige. This new instrument and the directions for administering were refined in a pilot study which included eight classes of sixth grade pupils in three different schools. The data from these classes were not included in the data of this present study.

The revised sociometric instrument which was then utilized for this study consisted of the following five questions:

A. Suppose you were to move to another classroom, which boy or girl from this classroom would you like best to go with you?

B. Suppose you were to move to another classroom, which boy or girl from this classroom would you not like to go with you?

C. Which boy or girl from this class would you like best to play outdoor games with you?

D. Which boy or girl from this class would you like best to be the president of your class?

E. Which boy or girl from this class would you like best to help you with your school work?

Selection of the Population

To test the hypothesis of this study and to determine the effect of certain recognized variables, the selection of the population was of prime importance. Obviously the population selected must have had the opportunity to acquire proficiency in reading since reading achievement was one factor in which status was to be determined. Status in peer prestige, the other factor, can be acquired only when pupils have had the opportunity to know each other in the types of situations to which they are expected to respond. To meet these two qualifications, sixth grade pupils were studied. Sixth grade pupils were studied because (1) elementary school pupils are kept together a greater part of the school day and have, therefore, a better opportunity to know each other and (2) the sixth grade class is the most advanced reading group in the elementary school. Primarily for purposes of convenience, the population was limited to sixth grade pupils in the Tulsa Public Schools.

In selecting the population, consideration also was given to one of the variables which might influence the relationship of reading achievement and peer prestige status, viz., the socio-economic level of the school community. Schools were chosen insofar as possible to be representative of the varying levels of socio-economic status with the full cognizance that some of the pupils attending any particular school would be representative of the adjacent socio-economic classes. The major point here is that the schools were located in a specific socio-economic community and that the value patterns of the majority of pupils would reflect the value patterns of that specific community. Consequently, four schools were selected in communities of "high" socio-economic level, three schools were selected in communities of "middle" socio-economic status, and eight schools in communities of "low" socio-economic status. This distribution accounted for eleven classes with 381 pupils from the "high", eight classes with 248 pupils from the "middle", and eleven classes with 352 pupils from the "low" socio-economic level. This totaled thirty classes and resulted in the acquisition of data on 981 pupils as the sample population of this study.

The Results

The sociometric questionnaire was administered in all the selected schools by the classroom teacher during May, 1957. In tabulating the responses, the three choices requested for each question were weighted as follows: three points for a first choice, two points for a second choice, and one point for a third choice. The choices listed for the social rejection question "B," were considered at all times as negative numbers. In order to provide the data for each of the components of peer prestige, the responses to each question were tabulated separately and the total peer prestige score for each person consisted of the totaled scores of the five questions. These scores, by classes, were arranged in rank order and categorized as nearly as possible into thirds. Thus a pupil's peer prestige status--high, middle, or low--was determined by the rank of his peer prestige score.

The reading achievement scores were also arranged by classes in rank order and categorized as nearly as possible into thirds. Thus a pupil's reading achievement status--accelerated, normal, or retarded--was determined by the relationship of his reading score to all the reading scores of the other members of only his class.

Relationship of Reading and Peer Prestige

The peer prestige status and reading achievement status of each pupil were plotted against each other into a three by three contingency table for each class. Then a summation of the frequencies of corresponding cells was made for all thirty classes and put into a summary contingency table. These data are presented in Table I.

To test the major hypothesis, the relationship of status in reading achievement and peer prestige, the chi-square technique was utilized for the entire sample population of 981 pupils. This chi-square value was 122.65, which is significant at or beyond the 5 per cent level. Consequently, on the basis of these data, the tentative conclusion that the relationship of peer prestige status and reading achievement status is significant appears justified.

In an effort to corroborate these findings, a second approach was used. Obviously, if a significant relationship exists between accelerated reading status and high prestige status and between retarded reading status and low prestige status, then an attempt to establish relationships between these extremes should also result in corroborating evidence for the major hypothesis. Consequently, this approach utilized only those pupils whose scores deviated most toward the extremes of peer prestige status--the stars, the rejectees, the isolates--and of those whose reading achievement scores deviated most from the mean of the class--the highly accelerated and the extremely retarded.

For the purposes of this study, stars were defined as the four persons who received the highest peer prestige scores in each class. Of the 120 stars, 79 were accelerated readers and only eight were retarded readers. The rejectees were defined as those four persons who received the lowest peer prestige scores in each class. Of these 120, only 17 attained accelerated reading achievement while 63 were classified as possessing retarded reading achievement status.

The isolates of this study were those pupils who were neither chosen nor rejected by anyone on any of the components which together make up the peer prestige score. Only fourteen of the thirty classes had isolates and the maximum number in any class was two pupils. The numbers in this case were insufficient to support any generalization but some significance might be attached to the fact that of the 20 isolates discovered, 13 were retarded in reading, five were accelerated, and only two were in the normal range.

Turning from peer prestige status to reading achievement status, the findings are similar. The highly accelerated readers were defined as those persons with the four highest reading achievement scores in each class. Of these 120 pupils, 77 had high peer prestige status and 16 had low peer prestige status. Of the 120 retarded readers, defined as persons with the four lowest reading achievement scores in each class, 71 also had low peer prestige status while only ten had high peer prestige status with their classmates.

The chi-square values of the relationship of status in reading achievement and peer prestige of the deviates as presented in Table II would seem to lend weight to the earlier conclusion of the significance of this relationship for the entire sample population. Consequently, on the basis of this population, a significant relationship exists between status in reading achievement and peer prestige.

The Effect of the Area of Peer Prestige

Inasmuch as a significant relationship between reading achievement status and peer prestige status has been established, the effect of the area in which a person has gained prestige with his peers is of considerable importance. Since ability in reading might be important in gaining prestige in the academic area, a highly significant relationship would be anticipated. Conversely, achievement in reading certainly would not be an integral factor in the areas of sports or social relationships.

In general, these common observations were

TABLE I

PEER PRESTIGE STATUS AND READING ACHIEVEMENT STATUS OF 981 PUPILS IN THIRTY CLASSES

Peer Prestige Status	Reading Achievement Status			
	A	N	R	n_r
High	159	111	49	319
Middle	94	136	113	343
Low	66	96	157	319
n_c	319	343	319	981

TABLE II

CHI-SQUARE VALUES OF DEVIATES

Type of Deviate	x^2
Stars	56.21*
Rejectees	22.69
Isolates	52.06
Highly Accelerated Readers	46.69
Extremely Retarded Readers	40.13

*Any value of 5.99 or greater is significant at or beyond 5 per cent level.

TABLE III

CHI-SQUARE VALUES OF VARIOUS AREAS OF PEER PRESTIGE STATUS RELATED TO READING ACHIEVEMENT STATUS

Area of Status	x^2
Peer Prestige	122.65*
Social Acceptability	62.48
Sports Leadership	19.62
Political Leadership	116.07
Academic Leadership	162.41

*Any value of 9.49 or greater is significant at or beyond the 5 per cent level.

TABLE IV

PEER PRESTIGE STATUS AND READING ACHIEVEMENT STATUS IN COMMUNITIES OF VARYING SOCIO-ECONOMIC LEVELS

		Areas of Peer Prestige Status			
Socio-Economic Level	Peer Prestige Status	Social Acceptability Status	Sports Leadership Status	Political Leadership Status	Academic Leadership Status
High	43.04*	35.37	16.55	46.77	57.30
Middle	38.51	29.28	13.08	46.32	48.90
Low	29.40	8.20	0.89	30.17	59.87

*Any value of 9.49 or greater is significant at or beyond 5 per cent level.

supported by the data although the statistical manipulation did not permit differentiating in degrees among the areas to the extent desired. An inspection of Table III indicates that the relationship between status in reading achievement and each of the areas of peer prestige is significant at or beyond the 5 per cent level of confidence. That the value expressing the relationship between academic leadership and reading achievement is extremely high in comparison with the value expressing relationship between sports leadership and reading is apparent. However, both are statistically significant. Guarding against the danger of overgeneralizing upon the basis of present data, the only conclusion possible is that the size of the chi-square value of each relationship varies according to the question asked.

The Effect of Socio-Economic Level

Some research workers (1, 12, 16, 17) have found that high aspirations are more common, and accelerated academic achievement is more valued, in high and middle socio-economic communities than in communities of a low socio-economic level. Therefore, the factor of socio-economic level of the school community was included as the second variable of the minor hypothesis. Or, to pose the problem in another way, is the relationship between status in reading achievement and peer prestige different for communities of different socio-economic levels? And, in an effort to refine the data further, the factor of socio-economic level was also considered as it affected each area of peer prestige.

The chi-square values for the relationship of peer prestige status and reading achievement status for each of the socio-economic levels are presented in Table IV. Also presented in this table are the chi-square values of the relationship of reading achievement status and status in each of the areas of peer prestige.

While the degree of relationship between peer prestige status and reading achievement status varied somewhat according to the socio-economic level of the school community, the chi-square values for these relationships indicated that all were statistically significant at or beyond the 5 per cent level of confidence.

However, an examination of the chi-square values which express the relationships of reading achievement status to status in the various components of peer prestige, according to the different socio-economic levels, reveals at least several instances where this relationship is not significant. While the relationship of reading achievement status to status in the different components of peer prestige varies considerably for the schools in the high and middle socio-economic level, all are consistently significant at or beyond the 5 per cent level of confidence. The relationship of status in reading achievement and each of the different areas of peer prestige in the low socio-economic level indicates a pattern which differs from the other two classes. The relationship of reading achievement status and social acceptability status and the relationship of reading achievement status and sports leadership status are both statistically insignificant. Obviously, the ability to excel in reading is not as important as other possible factors in being socially acceptable or in gaining leadership status in the area of sports in the lower socio-economic communities. In contrast, achievement in reading was recognized and accorded a place of importance in the areas of political and academic leadership in communities of low socio-economic level.

The relationship of reading achievement status and peer prestige status was statistically significant for all three socio-economic levels. In addition, the relationships of status in reading achievement and each of the various components of peer prestige according to the different socio-economic levels were all statistically significant except for the areas of social acceptability and sports leadership in the lower socio-economic level. These two were statistically insignificant.

Summary and Conclusions

The evidence presented in this study seems to justify the following conclusions:

1. A significant relationship exists between peer prestige status and reading achievement status.
2. This relationship holds true irrespective of (a) the type of peer prestige, or (b) the socio-economic level of the school community.
3. The relationship between status in reading achievement and the various components of peer prestige seems to be influenced by the socio-economic level of the school community.

A study of this type provides only tentative answers to a very few questions while uncovering a number of additional problems. Certainly the possibility that an excellent reader in a low socio-economic area is recognized by his peers as possessing a special type of leadership which is not transferable to other areas of leadership contrasts rather vividly with the apparent high correlation between reading achievement and sports and social acceptability in the other two socio-economic levels. Among the isolates, the discovery that five accelerated readers were found as well as thirteen retarded readers suggests that the deviate reader in both directions is most likely to be a social isolate.

Another important question is the constancy of the relationships discovered. Would the readministering of a similar sociometric instrument in one, three, or four years from the dates of this study result in findings which are similar? Or, to examine one of the most important and basic questions, does a cause-and-effect relationship exist between reading and peer prestige? Is retardation in reading a cause for low peer prestige or vice versa? Or is there a third alternative which provides an explanation, viz., reading achievement status and peer prestige status are two factors among a number of factors which compose a matrix of personality and intellectual characteristics which are typical of specific types of individuals?

BIBLIOGRAPHY

1. Bedoian, V. H., Sociometric Status and Mental Health of Overage and Underage Pupils in the Sixth Grade, Doctoral Thesis, Los Angeles: University of Southern California, 1950.
2. Bonney, Merle E., "A Study of the Sociometric Process Among Sixth Grade Children," Journal of Educational Psychology, XXXVII, 1946, 359-72.
3. Bonney, Merle E., "Values of Sociometric studies in the Classroom," Sociometry, VI, 1943, 251-54.
4. Jennings, Helen Hall, "Structure of Leadership: Development of the Sphere of Influence," Sociometry, I, 1937, 99-143.
5. Jennings, Helen Hall, Leadership and Isolation, New York: Longmans, Green and Company, 1950.
6. Kelley, Truman L., et al., The Stanford Achievement Test, Intermediate Battery Partial, Form JM, New York: The World Book Company, 1953.
7. Kuhlen, Raymond G. and Lee, Beatrice, "Personality Characteristics and Social Acceptability in Adolescence," Journal of Educational Psychology, XXXIV, September, 1943, 321-40.
8. Ladd, M., "Relation of Social, Economic, and Personal Characteristics of Reading Ability," Teachers College Contributions to Education, Number 582, New York: Columbia University, 1933.
9. McClelland, F. M. and Ratliff, John A., "The Use of Sociometry as an aid in Promoting Social Adjustment in a Ninth Grade Home-Room," Sociometry, X, 1947, 147-53.
10. Meyer, George, "Some Relationships Between Rorschach Scores in Kindergarten and Reading in the Primary Grades," Journal of Projective Techniques, XVII, December 1953, 414-25.
11. Moreno, Jacob L., Who Shall Survive?, Washington, D. C.: The Nervous and Mental Disease Publishing Company, 1934.
12. Murfin, Mark, "A Study of the Expressed Attitudes of Children from Two Socio-Economic Levels and the Relationships to Intelligence and Reading Achievement," Studies in Reading, IV, Bloomington, Indiana: School of Education, India University, 1952, 239-43.
13. Northway, Mary L., "Outsiders: A Study of the Personality Factors of Children Least Acceptable to their Age-Mates," Sociometry, VII, 1944, 10-25.
14. Northway, Mary L., A Primer of Sociometry, Toronto: University of Toronto Press, 1952.
15. Northway, Mary L. and Widgor, Blossom T., "Rorschach Patterns Related to the Sociometric Status of School Children," Sociometry, X, May, 1947, 186-99.
16. Sewell, William H., "Social Status and Educational and Occupational Aspirations," American Sociological Review, XXII, February, 1957, 67-73.
17. Sheldon, William D. and Carrillo, Lawrence, "Relation of Parents, Home, and Certain Developmental Characteristics to Children's Reading Ability," Elementary School Journal, LII, January, 1952, 265-69.
18. Solomon, Ruth H., "Personality Adjustment to Reading Success and Failure," Clinical Studies in Reading, II, Supplemental Educational Monographs, 77, Chicago: University of Chicago Press, 1953, 64-82.
19. Spache, George D., "Personality Patterns of Retarded Readers," Journal of Educational Research, L, February, 1957, 461-69.
20. Young, L. L. "Sociometric and Related Techniques for Appraising Social Status in an Elementary School," Sociometry, X, 1947, 168-77.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

DEVEREUX READABILITY INDEX

EDGAR A. SMITH
The Devereux Foundation
Devon, Pennsylvania

READABILITY REFERS to the "quality of a piece of reading matter that makes it interesting and understandable to those for whom it is written" (Good, 1945). In many respects, it is a parallel term to the more familiar "reading ability". The former refers to a characteristic of the book, the latter to a characteristic of the reader.

The readability of a book is determined, at least in part, by the words used and the sentences employed. Books that are easy to read contain simple words in short sentences. Difficult books usually are written in long, involved sentences and these sentences contain more difficult words. For many years people such as Flesch (1948), Lorge (1939), Spache (1954) and Dale and Chall (1948) have been working in this area. While each has developed his own unique method of estimating the readability of books, each includes these two factors (word difficulty and sentence length). The formulae differ primarily in the methods of determining the word difficulty factor. Dale and Chall (1948) use a list of words that are most common in our language. These words make up a large portion of the total words in books which are easily read, a smaller portion of more difficult books. Flesch (1948) uses a count of morphemes or a count of polysyllabic words. The method investigated in this study is similar in effect, but uses the average number of letters in a word. The primary purpose is to develop a formula that requires less time without undue loss of accuracy.

Most common methods of dealing with sentences have been a measure of the proportion of complex and compound sentences and a measure of the average number of words in a sentence. These two measures are related, since complex sentences contain, on the average, more words than simple sentences.

It will be noted that, while the definition of readability includes both interest and difficulty as factors in readability, the formulae emphasize difficulty. "Interest" is a very individual matter. While attempts have been made to include some interest factors such as reference to persons (Flesch, 1943) in a readability formula, they have not been particularly fruitful. It would appear that this is best left to the reader and to the professional judgment of the teacher or librarian.

Development of Formula

In the Devereux formula, the average number of character-spaces (letters, numbers, punctuation marks) in a word is used as the estimate of word difficulty. This is faster than using a word list. It is simpler than a syllable count in that there is seldom any doubt concerning the number of letters in a word. This is especially true for people not particularly versatile in English and grammar, as syllables and morphemes are at times ambiguous.

Validation of Measure of Word Difficulty

The assumption that word length may be substituted for word lists and syllable counts to determine the difficulty of words was tested in the following ways.

To test the relationship between word length and lists of commonly used words, a sample of ten words in each of four categories was taken from the Thorndike Word List (1944). These categories were words occurring more than 100 times, words occurring 50-99 times, words occurring 10-49 times, and words occurring less than 10 times per million words. The respective means were 4.4, 6.5, 8.5 and 8.1. F and t tests between groups 1 and 2 and between groups 2 and 3 were significant at the .001 level. No significant difference was found between the two most uncommon groups, 3 and 4. This would indicate that, up to a point, mean word length increased as the words became more uncommon.

There are 1069 words that occur more than 100 times per million, and 952 words occurring 50 to 99 times. Thus it appears that the thousand most frequently used words are the shortest. The next thousand are somewhat longer, yet shorter than words used less than 50 times per million.

To determine whether words that contain more syllables also tend to contain more letters, samples of words were selected at random from Webster's New International Dictionary (1939) and tested by means of the F test. Ten words of each of five classes were selected, the classes being words of one, two, three, four and five syllables respectively. The respective means were 4.4, 6.7, 9.0, 10.0, and 12.0. F test and t_{12} and t_{23} were significant at the .001 level and t_{45} at the .01 level. The t_{34} test was not significant but the difference was in the expected direction. It would appear that there is a significant increase in the number of letters, which accompanies the increase in the number of syllables.

Derivation of Formula Weights

Seventeen graded readers were examined to determine the average word length and average sentence length of each. These included readers from the American Book Company, the Betts-Welch series of readers, and texts from Scott-Foresman, covering grades two through eight. From these data, Pierson correlations were calculated between word length, sentence length and assigned grade level, and these correlations used to derive a regression equation. This is presented in Table I. In computing the multiple correlation (R = .74) the following regression equation was determined:

$$\text{Grade Placement} = 1.56\,W_L + 0.19\,S_L - 6.49.$$

It appears that this equation accounts for about half of the variance and compares favorably with the multiple correlation coefficients reported for the other formulae; Dale .70, Lorge .66, and Flesch .66 (Dale, 1948).

To simplify the computations contained in the equation, it was reduced to $8\,W_L + S_L$. To conserve time, the average number of spaces per line of text is used in calculating the average word length. This results in the addition of, and multiplication by constants, changing the absolute, but not relative values obtained.

These two simplifications do, however, result in numbers other than grade equivalents. Although this is a disadvantage, there are secondary advantages. The use of grade levels would often be damaging from a therapeutic standpoint when dealing with children in a special educational situation.

Validity of Index

Relation to Standard Reading Series. The Reading for Meaning (Guiler and Coleman, 1955) series provides very appropriate material for a validity check of the Formula. Experience with these booklets has indicated that they are accurately graded and present ample material for deriving a readability index. A single paragraph is not usually sufficient in length to obtain reliable word and sentence length measures. Since each booklet in the Reading for Meaning series contains many reading paragraphs, it presents ample material. The results are presented in Figure I. It will be noted that the readability scores progress from low to high with no inversions. The correlation between publishers grade level and readability scores would thus be virtually perfect ... Rho = 1.00.

While the formula was not derived from this Reading for Meaning series, it did form the major basis for establishing the Readability Index equivalents for the various functional reading grade levels. The rationale for this was based primarily upon prior experience with the series. Over a period of years it has stood the test of practical use. Since teaching staffs are familiar with it, it meets the qualifications of a good "reference yardstick", i.e., familiarity, practicality and statistical simplicity. Some statistical smoothing adjustment was required, but this was minimal.

Relation to Standard Achievement Test

The interpretation of written material subtest in the California Achievement Tests (Tiegs and Clark, 1957) presents additional material for verifying the validity of a readability formula. It is to be assumed that the material contained in the paragraphs to be read by fourth grade pupils would be more difficult to read than that presented to second grade pupils. Thus, the readability index should increase with progression from the Upper Primary through the Advanced level of the test. The Lower Primary level cannot be evaluated, since it does not contain equivalent material. The readability scores obtained from this material are presented in Table II. Again, it will be noted that the indexes increase from level to level. By reference to Table IV, it will also be noted that the grade levels obtained are consistent with those recommended by the test publisher, i.e., The Readability Index of 49.8 on the Upper Primary would place it in the third grade level, while the publishers recommend that the test be used with students in the third and lower fourth grades.

Reliability of Index

To estimate the reliability, the Readability Index was determined for two five page samples from each of thirty-five books. The Spearman-Brown formula was used to estimate the reliability of ten page samples. This was interpreted as a test-retest method, rather than a split-half method, since temporary factors were not operating to increase the correlations spuriously. The obtained estimates of reliability are reported in Table III.

Method for Determining Readability Index[1]

The following procedures were developed for

---

[1] To obtain a more detailed account of the development of the Devereux Readability Index, including instructions for computing, sample data sheet, and the indexes of 145 books, write to the Library, Devereux Foundation, Devon, Pennsylvania

FIGURE I

READABILITY INDEXES OBTAINED FOR READING FOR MEANING SERIES

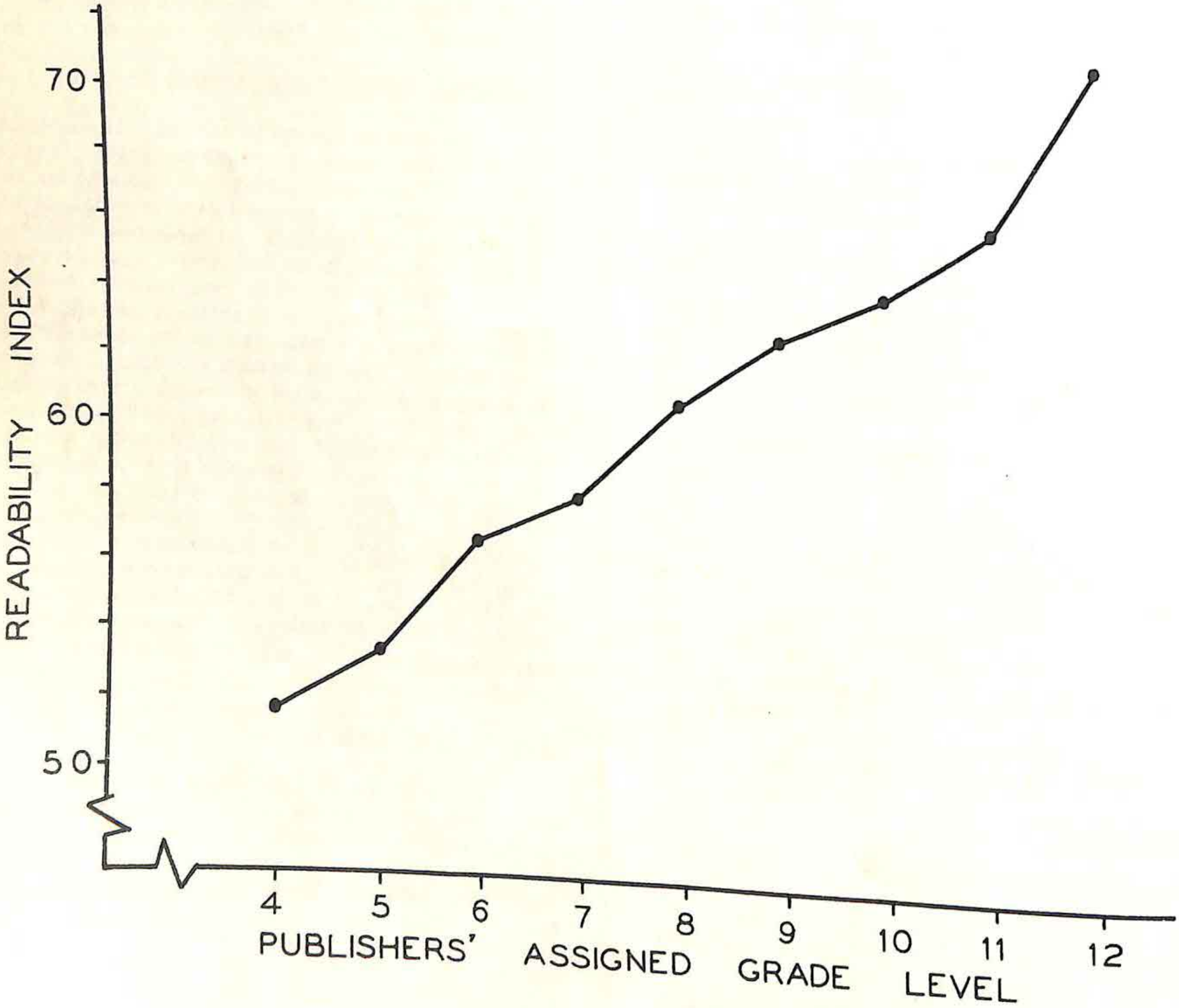


TABLE I

CORRELATIONS BETWEEN PUBLISHER'S GRADE LEVEL, WORD LENGTH AND SENTENCE LENGTH

	Grade	Word Length	Mean	s.d.
Grade			4.5	2.09
Word Length	.57		5.3	.43
Sentence Length	.69	.46	14.4	5.89

TABLE II

READABILITY INDEXES FOR READING INTERPRETATION PORTION OF THE CALIFORNIA ACHIEVEMENT TEST

Test Level	Word Length	Sentence Length	Total Index	Estimated Reading Levels[a]
Upper Primary Grade 3.0-4.5	5.25	7.86	49.86	3.1
Elementary Grade 4.0-6.9	5.64	14.89	59.97	7.7
Junior High Grade 7.0-9.9	5.80	17.98	64.35	9.7
Advanced Grade 9.0-14.9	6.22	18.82	68.62	11.5

[a] Obtained by interpolating between values in Table IV.

TABLE III

RELIABILITY ESTIMATES OF DEVEREUX READABILITY INDEX

	$r_{\frac{1}{2}}$	r	Mean		s. d.	
			odd	even	odd	even
Word Length	.62	.76	5.50	5.57	.204	.274
Sentence Length	.90	.95	14.4	14.7	4.35	4.37
Total	.89	.94	58.4	59.1	5.26	5.59

$r_{\frac{1}{2}}$ = Pierson r between 5 odd and 5 even samples.
r = r corrected by Spearman-Brown formula.

TABLE IV

CORRESPONDENCE BETWEEN GRADE LEVELS AND READABILITY INDEXES

Recommended Reading Level	Readability Index
2	47.4
3	49.6
4	51.8
5	54.0
6	56.3
7	58.5
8	60.7
9	62.9
10	65.2
11	67.4
12	69.6

Note - If the recommended functional reading level is the fourth grade, encourage the reading of books with a Readability Index of 51.8 or below.

computing the readability index:

1. The average number of character-spaces per line of text is established by selecting twenty lines on various pages in the book. The number of character-spaces in each line are counted (letters, punctuation marks, spaces between words, etc.). Care is taken to avoid the first and last lines of a paragraph, lines containing long blank spaces, or lines otherwise obviously atypical. The results are totaled and divided by 20 (carried to two decimal places and rounded to one) to obtain the mean number of character-spaces per line. The result is usually a number between 40 and 70.

2. The first of ten samples is selected. Samples should begin and end with a paragraph and should be approximately a page in length. Pages containing large pictures, tables or other interruptions should be avoided.

3. The number of character-spaces in the sample is estimated by multiplying the number of lines in the sample by the average number of character-spaces per line as determined in Step 1 and subtracting the number of blank spaces at the beginning and end of each paragraph.

4. If the sample includes titles or other headings, these should not be counted in estimating the number of spaces, the number of words or the number of sentences.

5. The total number of words in the sample is then counted. Hyphenated words are counted as one word and such numbers as "1492" as one word.

6. The number of sentences in the sample is then counted. Compound and complex sentences should be counted as a single sentence.

7. The above operations 2, 3, 5 and 6, are repeated for ten pages selected at varying places throughout the book, using approximately five odd and five even pages.

8. The total number of characters, total number of words, and total number of sentences in the ten samples is then computed.

9. The total number of characters is divided by the total number of words to obtain the average word length. Division should be carried to the fourth place and rounded to three.

10. The total number of words is divided by the total number of sentences to obtain the average sentence length. Division should be carried to the fourth place and rounded to three.

11. The values from steps 8 and 9 are entered in the formula:

$$8\,W_L + S_L = R.$$

(W_L stands for average word length, S_L stands for average sentence length, and R for readability index). The result may now be rounded to two decimal places.

12. Translation into grade level can be made by using Table IV, which is based on combined and smoothed data.

Summary

The Devereux Readability Index was developed from two factors (word difficulty and sentence length) found to be most frequent in a review of the literature. The method of determining word difficulty was simplified by using the average number of characters per word. This factor was verified by its relation to polysyllabic words and to lists of commonly used words. The average number of words per sentence was used as the measure of sentence length.

Weights for each factor were established by an analysis of graded readers and subject matter texts. The resulting formula was simplified as an aid to computation. It appeared to account for about one half of the total variance as do previous formulae.

This formula was verified by application to standard achievement tests and graded reading work books. A conversion table for translating from readability indexes to achievement test grade levels was presented. The reliabilities of each factor and for the total index were determined by an independent analysis of books and appeared to be adequate.

Step by step instructions for computing the readability of a book are presented.

REFERENCES

1. Dale, E. and Chall, Jean S. Formula for Predicting Readability. Ohio State University Educational Research Bulletin, 1948, 27, 11-20.
2. Flesch, R. Improved Readability Formulas. Journal of Applied Psychology, 1948, 32, 221-233.
3. Flesch, R. Marks of Readable Style, New York: Bureau of Publications, Teachers College, Columbia University, 1943.
4. Good, C. V. Dictionary of Education, New York: McGraw-Hill, 1945.
5. Guiler, W. S. and Coleman, J. H. Reading for Meaning, Chicago, Ill. :J. B. Lippincott, 1955.
6. Lorge, I. Predicting Reading Difficulty of Selections for Children. Elementary English Review, 1939, 16, 229-233.
7. McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, 1949.
8. Spache, G. Good Books for Poor Readers. George Spache, 1954.
9. Thorndike, E. L. and Lorge, I. The Teacher's Word Book of 30,000 Words, New York: Teacher's College, Columbia University, 1944.
10. Tiegs, E. W. and Clark, W. W. California Achievement Test (Form W) Los Angeles, California: California Test Bureau, 1957.
11. Webster's New International Dictionary. Springfield, Massachusetts: G. and C. Merriam, 1939.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

RELATIONSHIPS BETWEEN THE LEE-CLARK READING READINESS TEST AND THE 1937 REVISION OF THE STANFORD-BINET INTELLIGENCE TEST, FORM L

K. M. PARSLEY, Jr.
University of Alabama
MARVIN POWELL
Northern Illinois University

A NUMBER OF tests on the market purport to measure the child's readiness for reading upon entrance into the first grade of school. Generally, these tests are structured on the assumption that the "best" time to introduce reading is when the child has reached a mental age of about six years and six months. The scores of such tests indicate such factors as expectation of success in reading and/or the amount of delay indicated before the child will be ready for reading. Since there is stress on the concept of mental age and thus, indirectly, on intelligence, the authors felt there would be some value in comparing the results of a group of children on one such test, the Lee-Clark, with the results of the same children on the Stanford-Binet Test of Intelligence (1937 Revision, Form L).

Particularly, it was felt that the comparison of these tests would be most valuable for these children who seem "least" ready for reading when they enter first grade, since those who are ready are more easily recognized by teachers without testing. For the "poorer" child, the test used is often the only indicator which is available to the teacher of the probable amount of delay, although she may generally be able to recognize the greater immaturity of such children. It would generally be expected that those less ready for reading are also intellectually less mature. Thus, a reading test score might be expected to correlate well with an intelligence test score.

The problem of such relationships have been previously investigated—Anderson, et al. (1), Karlen (4), Moreau (5), Steinbach (6) – but generally the results of individual intelligence tests have not been used, except by Russell in 1956, and Burkes (3). The studies that have been done have been structured to include the entire range of readiness scores and intelligence test results, while this study employed a relative large, but select sample of children, all of whom had delay in readiness indicated. These are the children with whom the teacher has the greatest difficulty in reading instruction, and thus, additional information about this type of child should be valuable.

The Lee-Clark Test, a nationally standardized test which has been in use over the past 29 years, through three revisions, was used to determine reading readiness. According to the authors, it is most helpful in identifying children who are at the various stages of preparedness for beginning reading. The reliability studies indicate a split-half reliability of .93 for the total test and the validity studies indicate .56 and .65 correlation with two group intelligence tests. It was felt that this test would be representative of the reading readiness tests on the market and in common use.

The sample was composed of 169 entering first grade children in a suburban school district near Cleveland, Ohio. These children all had Lee-Clark Readiness Test grade placement scores of .7 or below, indicating a 3 month to 12 month delay. Each of these children was tested individually on Form L of the 1937 Revision of the Stanford-Binet Test of Intelligence by a competent, certified school psychologist. These children are from a generally middle-class socio-economic group, and are part of a first grade population whose average I.Q. is 112. The average chronological ages, mental ages, and I.Q.'s of the two groups studied are presented in Table I.

TABLE I

STANFORD-BINET AVERAGE CA, MA, AND I.Q.

Stanford-Binet	Lee-Clark Grade Placement	
	(.1-.7) N=169	(0-0) N=52
Average CA (Mos.)	78	78
Average MA (Mos.)	73	68
Average I.Q.	92.7	86.9

The low group, .7 and below, was selected primarily because teachers are usually concerned about the predictive values of the test at this level, while they are generally satisfied with the prediction above this point. Teachers generally assume that indicated delay is specifically related to mental immaturity, but evidence from other studies makes one uncertain of such a generalization. The children involved in this study were tested individually on the Stanford-Binet in order to determine whether or not this assumption could legitimately be made. The authors felt that the Lee-Clark was not really a good indicator of general mental maturity at the lower levels. Previous experience had led the authors to believe that the teachers found the I. Q. results of the children with low Lee-Clark scores of great value in placing the children in appropriate reading groups, more accurately than could be done on the basis of the Lee-Clark itself.

The major analyses involved determining the relationships between Lee-Clark raw scores and Stanford-Binet I. Q., Lee-Clark raw scores and Stanford -Binet mental age, Lee-Clark grade placement scores and Stanford-Binet I. Q., and Lee-Clark grade placement scores and Stanford-Binet mental age on the entire sample of 169 cases.

A secondary analysis was done on 52 cases in this total sample who had grade placements in the Lee-Clark of 0.0.* For these cases the raw scores on the Lee-Clark were correlated with the Stanford -Binet I. Q. and with the Stanford-Binet Mental ages.

Results and Discussion

The data used in the analyses were set up to make use of the raw score formula for the Pearson product-moment correlation technique (Blommer and Lindquist, 2) and are presented in the tables below and on the next page. The data presented in Tables II and III indicate the correlations between the variables studied for the total sample of 169.

The correlations between the Lee-Clark raw scores and the Stanford-Binet I. Q. and the Stanford -Binet Mental age are .35 and .39 respectively. These correlations are relatively low, indicating no strong relationship between these scores. Certainly one would hesitate to predict I. Q.'s or mental ages from Lee-Clark scores and vice versa.

The correlation between the Lee-Clark Grade Placement Scores and the Stanford-Binet I. Q. is .48 while that between the Lee-Clark Grade Placement Score and Stanford-Binet Mental Age is .43. These correlations are higher than those reported in the raw score comparison and offer better, but still relatively poor, prediction. However, the fact that all raw scores from 0 to 31 are assigned a grade placement score of 0.0 tends to obscure the lower end of the scale in this comparison and makes the correlation artificially high.

Therefore, it was decided to do an analysis of the 52 cases whose Lee-Clark grade placement score was 0.0. This data is presented in Tables IV and V. The correlations of the raw scores and the Lee -Clark and the Stanford-Binet I. Q. and the Stanford -Binet Mental Age were .34 and .25 respectively, indicating low positive relationships between these variables.

Generally then, these quantitative analyses show positive but low relationships between the Lee-Clark and the Stanford-Binet Tests. A cursory analysis of the data seemed to indicate some relationship might have existed. The data presented in Table I would indicate that the entire sample is five months retarded in M. A. as related to C. A. and the sample of 52 (0.0 on the Lee-Clark) was retarded ten months in M. A. as related to C. A. This seemed to relate to the Lee-Clark predictions of three to ten month delay for the .1 to .7 group, and a 12 month or more delay for the 0.0 group. It is felt, however, that the quantitative analysis is far more meaningful and, in a sense, negates the value of these figures. At best, it can only be said that those with Lee-Clark grade placements between .1 and .7 are retarded in I. Q. as related to a general population, and that those with Lee-Clark grade placements of 0.0 are more retarded than those with .1 to .7 grade placements.

Summary

The purpose of this study was to examine the relationships between the Lee-Clark Reading Readiness and the Stanford-Binet Test of Intelligence. An analysis of 169 cases with grade placements of .7 or below yielded low though positive correlations between the two tests. Specifically, the relationships between Lee-Clark raw scores and I.Q., Lee -Clark raw scores and mental age, Lee-Clark grade placements and I. Q., and Lee-Clark grade placement and mental age were .35, .39, .48 and .43 respectively. The latter two correlations are artificially high due to the grouping of all raw scores from 0 to 31 into a single grade placement category 0.0.

An analysis of 52 cases with grade placements of 0.0 on the Lee-Clark also yielded low positive correlations. The relationships between Lee-Clark raw scores and I. Q. and mental age were .34 and .25 respectively.

Although the total group was some five months delayed in mental age as related to chronological age, and the smaller group was some ten months retarded, there was no way of adequately predicting readiness on the basis of the mental age analysis.

---

* A grade placement of 0.0 is assigned to all raw scores from 0 to 31.

TABLE II

THE SUMS AND SQUARES OF THE VARIOUS LEE-CLARK AND STANFORD-BINET SCORES FOR THE TOTAL SAMPLE (N = 169)

N = 169	Lee-Clark Data		Stanford-Binet (1937) (Form L)	
	Raw Scores	Grade Placement	I. Q.	M. A.
Sum of Scores	5,746	46.1	15,667	12,290
Sum of (Scores)^2	207,430	2,125.21	245,454,889	151,044,100

Sum of (L-C Raw Scores) (S-B; I. Q.) = 538,828
Sum of (L-C Raw Scores) (S-B; M. A.) = 423,060
Sum of (L-C Gr. Place.) (S-B; I. Q.) = 4,318.3
Sum of (L-C Gr. Place.) (S-B; M. A.) = 173.6

TABLE III

PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN THE VARIABLES USING TOTAL SAMPLE (N = 169)

	Lee-Clark	
	Raw Score	Grade Placement
Stanford-Binet I. Q.	.35	.48
Stanford-Binet M. A.	.39	.43

TABLE IV

THE SUMS AND SQUARES OF THE VARIOUS LEE-CLARK AND STANFORD-BINET SCORES FOR THE SECONDARY SAMPLE (N = 52)

N = 52	Lee-Clark Data Raw Scores	Stanford-Binet (1937)(Form L)	
		I. Q.	M. A.
Sum of Scores	5,746	4,520	3,527
Sum of (Scores)^2	33,016,516	20,430,400	12,439,729

Sum of (L-C Raw Scores) (S-B; I. Q.) = 107,147
Sum of (L-C Raw Scores) (S-B; M. A.) = 83,425

TABLE V

PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN THE VARIABLES USING THE SECONDARY SAMPLE (N = 52)

	Lee-Clark Raw Scores
Stanford-Binet I. Q.	.34
Stanford-Binet M. A.	.25

REFERENCES

1. Anderson, I. H., Hughes, B. O., and Dixon, W. R. "Rate of Reading Development and Its Relationship to Age of Learning to Read, Sex, Intelligence," Journal of Educational Research, 1957, 50, 481-94.
2. Blommer, Paul, and Lindquist, E. F., Statistical Methods in Psychology and Education, Houghton Mifflin Company, Boston, 1960.
3. Burkes, H. F. and Bruce, Paul, "The Character of Poor and Good Readers as Disclosed by the Wechsler Intelligence Scale for Children," Journal of Educational Psychology, 1955, 46, 488-93.
4. Karlin, R., "The Prediction of Reading Success from Readiness Tests," Elementary English, 1957, 34, 320-22.
5. Moreau, Margaret C., "Long Term Prediction of Elementary Reading Achievement," California Journal of Educational Research, Vol. 1, No. 4, September, 1950.
6. Steinbach, Sister Mary Nila, "An Experimental Story of Progress in First Grade Reading," Educational Research Monograph, Catholic University of America, Vol. XII, No. 2, June 1, 1940.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

ORAL READING USED AS AN INDICATOR OF REACTIONS TO FRUSTRATION *

GLADYS NATCHEZ
City College of New York

THE RELATIONSHIP between reading disabilities and emotional problems has received concerted study and attention for many years. Emotional disturbance has been cited as both the cause and the result of reading disability (1). Few research studies, however, indicate satisfactorily how these factors function or in what way they affect each other. In order to investigate these relationships further, it is important to examine the child's characteristics while he is directly engaged in the reading task.

It is of especial importance, in this connection, to examine the child with reading disability because 1) a definite relationship has already been established between reading disability and emotional disturbance (2, 3), and 2) specific personality traits of these children have been reported (4, 5).

From related studies (6, 7, 8), it seems reasonable to conclude that whatever the original personality structure may have been, the child with reading disability manifests reactions related to frustration. The most common of these reactions have been described in the literature as dependence, aggression, and withdrawal. Such frustration reactions may become so intense that they interfere with the learning process, which in turn can lead to emotional maladjustment. For this reason, the investigator endeavored to determine whether the child with reading disability is in general characterized by a high proportion of frustration and how such frustration may interact with the reading situation.

It was hypothesized that 1) the retarded reader will display more frustration-type reactions in the classroom reading situation than the non-retarded reader; 2) the retarded reader will react to the reading situation in the way he reacts to frustrating situations in general. Specifically, the retarded reader who repeatedly reacts to frustration with dependence, aggression, or withdrawal, will show dependence, aggression, or withdrawal in his behavior during the reading situation.

Procedure

First, two groups of thirty retarded and non-retarded readers each, were matched on the basis of race[2], school class[3], grade[4], sex[5], age and intelligence. They differed with regard to reading ability. The extent to which the groups were similar in age and intelligence and different in regard to reading level appears in Table I.

TABLE I

COMPARISON OF I. Q., CHRONOLOGICAL AGE, AND READING LEVEL FOR RETARDED AND NON-RETARDED READERS

	Retarded Readers	Non-Retarded Readers
Mean I. Q.	100.63	101.37
Mean C. A.	10.60	10.66
Mean Reading Level	3.3	7.1

A study of the matched groups in the classroom reading situation was evolved to test whether the retarded readers would display more frustration in the classroom reading situation than the non-retarded readers. The class teacher was asked to conduct a regular reading period. As the subjects read, reactions were recorded by the investigator and a reading specialist on the Check Sheet to Record Frustration Reactions.[6]

Each teacher was familiar with the procedure to be used and had informed the class of the experiment. The class was cautioned to remain silent if someone had difficulty with a word. The pupils were asked not to call out words or to help the reader in any way. Five minutes were allowed for each subject to read a passage. When the child came to an unfamiliar word, he was given a maximum of thirty seconds to react. As soon as he asked for help, became upset, or showed any other definitive reaction, he was given assistance and continued reading. The reading material consisted of the book

---

*Footnotes will be found at the end of this article.

normally used when the class was engaged in a reading lesson.

The next procedure compared the retarded reader's reactions while reading, with his frustration reactions in general. In order to sample a sufficiently large number of subjects, ultimately to form relatively pure groups descriptive of the variables dependence, aggression, or withdrawal, fifty-nine retarded readers were selected. The retarded readers were chosen on the same basis as those in the matched groups. Each subject participated in the classroom reading situation just described and in addition participated in an individual or controlled equivalent reading situation. In the controlled reading situation, each child was called out of class individually and presented with a selected passage in the reader customarily assigned to his class. They were all given identical instructions. Reactions during this reading situation were also recorded on the Check Sheet for Recording Frustration Reactions.

These fifty-nine subjects were then examined with regard to their frustration reactions in situations other than reading. These situations yielded data gathered from school records, peer ratings, teachers' judgments, and a frustration-type task. Suitable instruments were devised to record dependence, aggression and withdrawal reactions manifested during these varied circumstances. Composite scores for each variable were then derived. These were next compared with composite reaction scores derived from the two reading situations.

Results and Conclusions

The children with reading disability manifested a significantly greater proportion of dependence, aggression, and withdrawal reactions while reading than did the children without such difficulty. Results can be seen in Table II.

TABLE II

DIFFERENCES BETWEEN RETARDED AND NON-RETARDED READERS' SCORES ON THE CHECK SHEET FOR RECORDING FRUSTRATION REACTIONS

	Retarded Readers	Non-Retarded Readers
Mean	7.60	.47
S.D.	3.02	.97

It can also be deduced from this table that retarded readers perceive the reading task as a frustration situation.

Secondly, a significant relationship existed between the retarded readers' reactions in the reading situation and their reactions in frustration situations other than reading. Comparison of data secured in situations other than reading with data descriptive of reactions to reading is shown in Table III. It can be seen from this table that all the like variables are in a positive direction, while all but one of the unlike variables are negative. On the basis of this analysis, it would seem justifiable to conclude that frustration-type behavior for the retarded reader in a reading situation can be predicted from what is known of his behavior to frustration in situations other than reading.

Further evidence to substantiate this conclusion was gathered when the subjects were classified into criterion groups. Out of the fifty-nine retarded readers, eleven were able to be classified as manifesting a predominance of dependence, thirteen exhibited a preponderance of aggression, while eleven utilized mainly withdrawal reactions. The frequency with which subjects from each of these "pure" groups could be reclassified in terms of dependence during reading is shown in Table IV.

Table V shows how those subjects whose behavior was first categorized as aggression could be reclassified as a result of observed reactions in a reading situation.

Subjects who could be reclassified into the withdrawal category on the basis of reactions during the reading situation are shown in Table VI at the top of the last page. It might be justifiably concluded from these data that children with reading disability will tend to exhibit in a reading situation, the same kind of behavior that they normally display in frustration-type situations.

Implications

The two major findings of this study lead to important implications. The first implication relates to the high degree of frustration found in the retarded readers. Such disproportion of detrimental characteristics, in turn, must lead to interference with the learning process. This finding suggests that school personnel in general and remedial specialists in particular concentrate their efforts on the prevention and decrease of such frustration throughout the school years. It seems of particular importance for pupils in the early grades to be free from frustration-producing situations so that emotional reactions connected with school work can be avoided as far as possible.

The second implication suggests that children with reading disability are in need of specialized instruction which combines appropriate techniques from the fields of both psychology and education in order to deal effectively with the decided frustration component found in these children. In other words, the child with reading difficulty, in all probability, carries with him feelings of defeat, anger, and bewilderment. Remedial procedures which ignore these interfering characteristics may prove futile

TABLE III

CORRELATION COEFFICIENTS BETWEEN COMPOSITE SCORES OF DEPENDENCE, AGGRESSION, AND WITHDRAWAL IN READING SITUATIONS COMPARED WITH COMPOSITE SCORES OF DEPENDENCE, AGGRESSION, AND WITHDRAWAL IN SITUATIONS OTHER THAN READING

	Schools		
	A	B	C
Between Like Variables			
Dependence-Dependence	.650*	.588*	.037
Aggression-Aggression	.745*	.624*	.728*
Withdrawal-Withdrawal	.577**	.721*	.679*
Between Unlike Variables			
Dependence-Aggression	-.085	-.469	-.407
Dependence-Withdrawal	-.351	-.055	-.092
Aggression-Dependence	-.265	-.275	.147
Aggression-Withdrawal	-.422	-.482	-.446
Withdrawal-Dependence	-.192	-.070	-.428
Withdrawal-Aggression	-.480	-.531	-.562

* Significant at the one percent level of confidence

** Significant at the five percent level of confidence

TABLE IV

TESTING THE CRITERION GROUPS FOR DEPENDENCE IN READING

	Dependence	Aggression	Withdrawal
Classifiable	5	1	0
Non-Classifiable	6	12	11

TABLE V

TESTING THE CRITERION GROUPS FOR AGGRESSION IN READING

	Dependence	Aggression	Withdrawal
Classifiable	0	8	0
Non-Classifiable	11	5	11

TABLE VI

TESTING THE CRITERION GROUPS FOR WITHDRAWAL IN READING

	Dependence	Aggression	Withdrawal
Classifiable	0	0	8
Non-Classifiable	11	13	3

and ineffective. An approach, on the other hand, that recognizes and treats the child's attitudes as well as his academic weaknesses, may result not only in the mastery of reading skills, but in the ultimate improvement of his emotional disturbance.

FOOTNOTES

1. Published in monograph form by New York University Press, December, 1959.
2. Only white children were selected.
3. For each retarded reader selected in a particular class, a non-retarded reader was selected from that same class.
4. The sample was limited to children in grades five and six.
5. The experiment was limited to boys.
6. The "Check Sheet for Recording Frustration Reactions" together with the accompanying manual can be found in Personality Patterns and Oral Reading, by Gladys Natchez, New York University Press, 1959, New York.

REFERENCES

1. Smith, N. B. "Reading in 1953," Education, LXXIII (May, 1953) pp. 532-58.
2. Odenwald, R. P. and Shea, J. A. "Emotional Problems of Maladjustment in Children With Reading Difficulty," American Journal of Psychiatry, CVII (1951), pp. 890-93.
3. Gann, E. Reading Difficulty and Personality Organization, New York: King's Crown Press, 1945.
4. Stauffer, R. "Certain Psychological Manifestations of Retarded Readers," Journal of Educational Research, XCI (February, 1948), pp. 435-52.
5. Spache, G. "Personality Characteristics of Retarded Readers as Measured by the Picture Frustration Study," Educational Psychological Measurements, XIV (1954), pp. 186-91.
6. Parker, I. "Personality Problems of Disability," National Elementary Principal, XIX (July, 1940), pp. 603-10.
7. Blanchard, P. "Psychoanalytic Contributions to Problems of Reading Disabilities," Psychoanalytic Study of the Child, II (1946).
8. Freud, A. and Burlingham, D. War and Children, New York: International University Press, 1943.
9. Dahlberg, C., Roswell, F., and Chall, J. "Pschotherapeutic Principles ad Applied to Remedial Reading," Elementary School Journal, LIII (December, 1952), pp. 211-17.
10. Maier, N. Frustration, New York: McGraw-Hill Book Co., 1949.
11. Roswell, F. "Observations on Causation and Treatment of Learning Disabilities," American Journal of Orthopsychiatry, No. 4, XXIV (October, 1954), pp. 784-88.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

WHAT UNDERSTANDINGS DO SECOND GRADE CHILDREN HAVE OF TIME RELATIONSHIPS?

J. D. McAULAY
Pennsylvania State University
University Park, Pennsylvania

DURING THE month of September, 1959, one hundred and sixty-five second grade children were tested on their understanding of time relationships. The study revolved about the question: have second grade children developed understandings of time which are (a) associated with self, (b) associated with the immediate environment, (c) associated with historical events. It was hoped a study of this question might indicate:

(a) Do second grade children understand the relationship of the past to the present?
(b) Are these children capable of understanding periods of time; can they pinpoint events in time which serve as marking places about which experiences might cluster or focus?
(c) Are these children capable of projecting themselves from the immediate present into past time?
(d) Do these children have an understanding of past social reality?
(e) Do these children have an understanding of the continuity of time?

By time relationships is meant continued duration, existence and progress affecting persons, things and events.

Eighty children from the study were selected from School A, serving a suburb of middle socio-economic families of the states largest industrial city; forty-two were selected from School B, children from lower socio-economic families whose livelihood centered about the railroad industry; forty-four children were selected from School C, serving a professional college community. Sixty-two third grade children from a consolidated school serving a farming community were also tested on their understanding of time relationships. It was hoped the study of these third grade children might give a comparison and trend of the development of time understandings with second grade children.

The range of I. Q. scores from the Otis Quick Scoring, given during the first week of the school year, 1959, for these children are shown at the top of the next page.

The mean chronological age of the 165 second graders was 7 years, 4 months. The mean chronological age of the 62 third graders was 8 years.

The children were given the following problems orally and individually and the responses recorded either by the child himself on answer sheets supplied or by the adult giving the question on a data recordung sheet.

The problems were divided into three groups:

A. Those problems concerning time concepts associated with self.
   1. Whom do you think has lived longer. Your mother or your grandmother (or grandfather)?
   2. Here on the blackboard draw a line showing how long (many years) you think your grandmother has lived. Now draw a line showing how long your mother has lived.
   3. Draw a line showing how old you are. Then draw a line showing how old your mother is.
   4. Write how old you are. Write how old your brother (or sister) is. Now draw a line showing how old you are. Draw a line under it showing how old your brother or sister is.
B. Time concept associated with the immediate environment.
   1. Which is longer -- a week or a month
      a month or a year
      a second or a minute
   2. Which will be here first, Christmas or Easter?
      What do you remember about Christmas?
      What do you remember about Easter?
   3. Which is longer, the summer vacation or the Christmas vacation?
   4. Who is older, your father or President

DISTRIBUTION OF I. Q. SCORES AMONG SCHOOLS

I. Q. Range	School A	School B	School C	3rd Grade
65 - 85	9	11	4	8
86 - 105	27	19	9	12
106 - 125	33	9	21	31
126 - 145	11	3	9	11
	80	42	43	62

Eisenhower?

5. Which is the longer time, between Halloween and Thanksgiving; or between Christmas and Easter?

C. Time concept associated with historical events.
   1. Who lived first, Washington or Lincoln?
      (a) Tell something about the time when Washington lived.
      (b) Tell something about the time when Lincoln lived.
   2. Who lived here first, the White people or the Indians?
      (a) Draw something about the long-ago White people who lived here.
      (b) Draw something about the Indians when they lived here long ago.
   3. Which was a state first, Pennsylvania or Alaska?
   4. How many years do you think--
      (a) Pennsylvania has been a state?
      (b) This town has been here?
      (c) This school has been here?

The problems were arranged and given in order of their power.

Summary of the Study

The study seemed to indicate that second grade children have a more clear understanding of the past than they do of the present; they seem to have a better understanding of the removed time environment than of the immediate and personal time environment.

(a) Second grade children have difficulty associating the past with the present in the immediate environment. There seems to be little understand- of the time relationship between child, mother, and grandmother (or grandfather). However, the second grade child has a clearer understanding of the time relationship between the father and the President of the United States, and an even fuller understanding of the time relationship of the present to the past, if such are completely disassociated from self. The time relationship of Indians to white pioneers seems to be understood by the majority of the second graders interviewed more clearly than the time relationship between mother and child.

These second grade children do seem to have an understanding of the past as related to the present if the time relationship is removed from self and the immediate environment.

(b) Second grade children seem to be capable of understanding periods of time. However, periods of time are more clearly understood if pivoted about events rather than persons or places. The child has a finer conception of experiences associated with Indians and the pioneers than with Washington and Lincoln, or with the school, town, and state. The child seems to be able to project himself into the past as regards persons, places and events, but probably more efficiently with events. He is capable, seemingly, of understanding the relation of himself to past events, whether it be the summer vacation recently completed, or the time of the Indians and pioneers.

(c) Second grade children do have some understanding of past social reality, if such is not related directly to the immediate concern of the child or his personal environment. He seems to have some understanding of the lives of Lincoln and Washington, of the pioneers and early Indians but little of the the background of his own mother and grandmother. He has collected a greater number of understandings about calendar events such as Christmas and Easter than he has about persons related to time sequence. Time events seem to be more efficient for the accumulation of information than persons or places.

(d) Second grade children seemingly have little understanding of the continuity of time, again particularly as related to the immediate and that closely associated with self. The child seems to have no understanding of the resumption and continuity of family from grandmother, mother to child. He has a better understanding of calendar continuity --- he knows the summer vacation is longer than the Christmas vacation, that the calendar period between Christmas and Easter is longer than that between Halloween and Thanksgiving. He has less understanding of time continuity as related to place than he has of calendar continuity. He is somewhat confused in the time continuity of his native state as compared to one of the more recent states. His understanding of the time sequence relationship of school to town and to state is not clear, although he seems to have some understanding that the school is first in the time sequence. He seems to see little continuity in time between the older location and the younger one.

The overall conclusion of this study would seem to be that the social studies curriculum for the second grade underestimates the childs understanding of time. Perhaps the seven year old is able to comprehend large events of the past as they relate to the present; he is capable of associating historical persons one to the other, and retain some information concerning each event and person. The child would seem to have little comprehension of time as it relates to the immediate family and community. However, it is about the family and community the second grade social studies curriculum revolves.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 8, April 1961)

THE ALIFERIS MUSIC ACHIEVEMENT TEST AS A PREDICTOR OF SUCCESS IN MUSIC THEORY AT ST. OLAF COLLEGE

ADOLPH WHITE
St. Olaf College, Northfield, Minnesota

COLLEGE FRESHMEN who enroll as music majors have a great deal at stake when they register for the beginning course in music theory. They commit themselves to study in a course which demands a specialized background and aptitude, and which usually involves earning from four to eight credits over the period of a semester or a year. The outcome of this investment in time and effort (and money) is often left in doubt, and the student can only hope that he will be successful.

The first attempts to provide some basis for predicting success in music dealt almost exclusively with aptitude tests. It was supposed that a measurement of a student's capacity to learn music was a good predictor of his subsequent success. Some difficulty was encountered, however, as aptitude tests were challenged as valid instruments in measuring musical talent. In addition, it seemed that another variable--the student's background in music--might be even more important to success than his basic aptitude.

Problem

The Aliferis Music Achievement Test is specifically designed to measure what the author calls "auditory-visual discrimination."[1] This ability to "see with the ears" and "hear with the eyes" is a recognized competence indispensable to musicianship. An important part of the total work in music theory is devoted to developing this ability. It seems to be a tenable premise, therefore, that those students who show a marked development in this ability will excel in music theory study, while those with limited ability in this area will be at a decided disadvantage. The extent to which this premise is true can be shown by correlating test results with grades earned in music theory.

A more complete picture of the worth of achievement tests in predicting success in music theory may be obtained by comparing the correlation results described above with those derived from correlating aptitude test scores and music theory grades. In addition, the standardized norms should be compared with the mean and standard deviation computed for the group to be tested.

Procedure

Data were compiled from a group of 59 St. Olaf College students who are members of the classes graduating in 1960 and 1961. For each of these students, the following information was available.

1. Aliferis test score at beginning of freshman year.
2. Seashore aptitude test score at beginning of freshman year.
3. Aliferis test score at end of sophomore year.
4. Grades in music theory for four semesters of study.

The music theory grades ranged from A to D, and were given a numerical value as follows: A = 3, B = 2, C = 1, D = 0. There were no F grades in the study since those students earning an F in music theory did not complete four semesters of work.

The students were tested on four parts of the 1956 revision of the Seashore Measures of Musical Talents.[2] Due to limitation in time, only the tests on pitch, loudness, rhythm and total memory were given.

Results

In comparing Aliferis test scores at the beginning of the freshman year with theory grade averages for four semesters of study, the coefficient of correlation was .63. In comparing Seashore test scores at the beginning of the freshman year with theory grade averages for four semesters of study, the following coefficients of correlation were found:

---

1. Aliferis Music Achievement Test Manual, University of Minnesota Press, Minneapolis, 1954.
2. Seashore Measures of Musical Talents, The Psychological Corporation, New York, 1956.

Pitch .30, Loudness .19, Rhythm .15, Tonal Memory .12. It can readily be seen that the Alferis Music Achievement Test is far superior to the Seashore Measures of Musical Talents as a predictor of success in music theory for the 59 students studied.

Since the 59 students involved comprise a rather select group, by virtue of their having completed two years of theory study, it is not surprising that the mean for this group is considerably higher than the derived scores mean reported in the Aliferis Manual. Other students in their respective classes, who scored low at the beginning of the freshman year, were subsequently eliminated before completing four semesters of theory study. Indeed, some were eliminated before beginning any theory study. The mean for the 59 students on the Aliferis test was 36.27, and the standard deviation 9.68. The derived scores mean reported in the Aliferis Manual for 1,768 students is 31.13, and the standard deviation 10.15. The total number of items in the Aliferis test is 64.

However, further study of all the students in the classes of 1960 and 1961 who took the Aliferis test reveals that the 59 students came from a comparatively average group. The mean for 147 students in the two classes who originally took the Aliferis test at the beginning of the freshman year was 32.45.

Further, a study of 369 St. Olaf College students who took the Aliferis test at the beginning of the freshman year over a period of five years revealed a mean of 31.62 and a standard deviation of 8.84. This mean compared very closely with the derived scores mean of 31.13 reported in the Aliferis Manual, while the standard deviation indicated that St. Olaf students' scores were not as widely dispersed as those reported in the Aliferis Manual.

Further evidence of validity for the Aliferis test as a measure of the students' achievement in music is provided by a second administration of the test at the end of the sophomore year, or after two years of theory study. In comparing Aliferis test scores at this point with the average theory grades for the 59 students involved, the coefficient of correlation was .77. This correlation shows an even higher degree of unanimity between the test results and the grades given by theory teachers. The mean for the 59 students after two years of theory study was 47.09.

The theory grades earned by the 59 students ranged from a C- or D average (two C's and two D's) to a straight A average for the four semesters of theory study. Students with lower averages did not complete four semesters of study and could not be included in this study. On the basis of numerical equivalents listed above, a straight A average is represented by a 12 (4 3's) a straight B average by an 8 (4 2's), a straight C average by a 4 (4 1's), and so on for all other possible combinations of grades. The mean grade for all 59 students was 7.29, or slightly under a B average. This comparatively high average may partially be accounted for by the fact that most students with poor grades did not finish four semesters of theory study.

Using the data given above, it was possible to construct a table giving the estimated average grade in music theory for each score on the Aliferis test. The formula for computing these predicted grades is as follows:

$$Y_i = \overline{Y} + r_{xy}\left(\frac{\sigma y}{\sigma x}\right)(X_i - \overline{X})$$

Where:

Y_i = Estimation of average theory grade from Aliferis test score

$\overline{Y}$ = Mean for theory grades

r_{xy} = Correlation coefficient for theory grades and test scores

σ_y = Standard deviation for theory grades

σ_x = Standard deviation for test scores

$\underline{X_i}$ = Test score

$\overline{X}$ = Mean for test scores

Then, for example, the predicted average theory grade for a test score of 40 is:

$$Y_i = 7.29 + .63\left(\frac{2.69}{9.68}\right)(40 - 36.27) = 7.94$$

Thus, a predicted average theory grade for a test score of 40 is approximately a B.

The standard error of estimate for this predicted grade is computed as follows:

$$SEE = \sigma_y \sqrt{1 - r^2_{xy}} = 2.69 \sqrt{1 - (.63)^2} = 2.10$$

Thus, we may conclude that for persons with a test score of 40, we would be right two times in three if we estimate their average theory grade between 5.8 and 10.0. A complete table for predicting average theory grades appears at the top of the next page (Table I).

Finally, for purposes of comparison, the means for the Seashore test scores are also reported. Since each portion of the Seashore test must be considered separately, Table II is provided.

Conclusion

Undoubtedly, one of the important factors contributing to the lack of effectiveness of the Seashore tests in predicting success in music is the tests' limitations in being truly selective or discriminative. For example, in the pitch test, the range of scores among the 59 students measured was from 39 to 50. Thus, the student with the lowest score was able to get 78% of the items correct, while a perfect score was possible for one student. An even more striking example may be found in the tonal memory test where the range of scores was from 24 to 30, with 22 of the 59 students scoring a perfect score of 30.

When used with other measures, the Aliferis Music Achievement Test is an important tool in predicting success in music theory study at St. Olaf College. These other measures include a piano

TABLE I

TABLE FOR PREDICTING AVERAGE THEORY GRADES FROM TEST SCORES

Aliferis Test Score	Average Theory Grade Predicted	Aliferis Test Score	Average Theory Grade Predicted
64	12.14	39	7.77
63	11.97	38	7.59
62	11.79	37	7.42
61	11.62	36	7.24
60	11.34	35	7.07
59	11.27	34	6.89
58	11.09	33	6.72
57	10.92	32	6.54
56	10.74	31	6.37
55	10.57	30	6.19
54	10.39	29	6.02
53	10.22	28	5.84
52	10.04	27	5.67
51	9.87	26	5.49
50	9.69	25	5.32
49	9.52	24	5.14
48	9.34	23	4.97
47	9.17	22	4.79
46	8.99	21	4.62
45	8.82	20	4.44
44	8.64	19	4.27
43	8.47	18	4.09
42	8.29	17	3.92
41	8.12	16	3.74
40	7.94	15	3.57

TABLE II

SEASHORE TEST SCORES AND MEANS

	Number of Items	Mean reported in Norms Table	Mean for 59 St. Olaf Students
Pitch	50	40.4	45.9
Loudness	50	42.8	45.1
Rhythm	30	26.5	28.2
Tonal Memory	30	25.0	28.7

test, the American Council on Education Psychological Examination for College Freshmen, and the rank in the student's high school graduating class. The last two measures, however, are more likely to be predictive of a student's success in any collegiate endeavor.

RESEARCH NEWS

JOHN SCHMID, Editor
University of Arkansas
Fayetteville, Arkansas

Readers are invited to send items of research interest to John Schmid, Professor, College of Education, University of Arkansas, Fayetteville, Arkansas.

NEW STATUS FOR SOCIAL STUDIES

The National Science Foundation, which is the Federal Goverment's Agency for reporting basic scientific research, has created a new division for research which includes antropological, economic, and sociological sciences. This new division will be headed by Henry W. Riecken, formerly with the University of Minnesota.

NEW YORK STATE CENTER FOR EDUCATIONAL RESEARCH

The State Commissioner of the New York State Education Department, James E. Allen, Jr., has proposed the development of a state center for education experiments to act as a clearing house for new ideas, to stimulate educators by challenging old ways, and to stir interest in new possibilities. It is hoped that this center will reduce the time lag between recognition of worthwhile experiments in education and their general adoption.

FEDERAL SUPPORT FOR COMMUNICATIONS RESEARCH

The Federal Government is supporting research which covers a wide variety of problem areas relating to educational uses of the newer media of communication. This program is administered by the Educational Media Branch in the Division of Statistics and Research Services of the Office of Education, U. S. Department of Health, Education, and Welfare. Grants are made to public or non-profit organizations. The program currently supports more than one hundred research projects. Recently, a small grant program has been initiated to encourage small scale exploratory studies and to promote active participation in media research by relatively inexperienced investigators. A limitation of $2500 in direct costs plus an allowably fifteen percent for indirect costs is established for these so-called small grants.

Individuals interested in further information or in receiving News and Reports should write to: Educational Media Branch, Office of Education, Department of Health, Education, and Welfare, Washington 25, D. C.

RESEARCH AT THE UNIVERSITY OF PENNSYLVANIA

Two studies are in progress in the area of reading: "Factors Contributing to Dyslexia" and "A Comparison of the Reading Achievement of German and American Children." Two studies which have just recently been completed are "Factors Associated with Progress in a Reading Improvement Course for College Students" and "Improving the Item Validity of Study Habit Inventories."

NEW PROGRAM AT THE UNIVERSITY OF ILLINOIS

Beginning in September, 1960, the University of Illinois was allotted five fellowships under Title IV of the National Defense Education Act, 1958. These graduate fellowships are in a newly instituted program, "Education-Psychology of Classroom Learning and Instruction." The principle aim of this new program is to train college instructors in this field of knowledge as well as research workers and staff specialists concerned with the psychological aspects of curriculum development in instructional improvement in public schools, military, and industrial settings. The program leads to a doctoral degree in education. Further information may be obtained from Professor L. M. Stolurow, Research Coordinator, Institute on Research for Exceptional Children, University of Illinois.

RESEARCH AT THE UNIVERSITY OF KANSAS

Some research dissertations under way in the School of Education at the University of Kansas are:

1. An experiment in Grouping for Arithmetic Instruction in the Sixth Grade.
2. The Effect of External Feedback on Vocal Production.
3. The Relation Between Intelligence, Music Ability, and Asetic Sensitivity.
4. A Music Achievement Test for Intermediate Grades.
5. A Comparison of Certification Regulations for Superintendents, Principals, Counselors, and Supervisors Established by Educators with Regulations Established by Lay Persons.
6. A Study of the Aspects of Differentiation of Abilities in Interpretation of Reading with a Group of Gifted Children.
7. A Differential Analysis of Some Intellective and Effective Characteristics Peer Accepted and Rejected Adolescent Children.

THE PREDICTIVE VALIDITY OF SCAT

Mr. Fred M. Smith at Louisiana State University and Agricultural and Mechanical College conducted a study on the predictive validity of the School and College Ability Test for undergraduate academic achievement. Three hundred and forty-eight male and female students were used. This study is interesting in that prediction of long range undergraduate achievement was investigated. The validity coefficients for the first, second, third, and fourth years respectively were .46, .35, .39, and .39.

EXPERIMENT IN FRENCH LANGUAGE INSTRUCTION

A two-year experiment was conducted at Antioch College on the use of new teaching procedures in French language instruction. The new program reorganized the teaching procedures so that a major portion of the classroom time previously conducted by the instructor was now handled by two student laboratory assistants. These assistants worked from a series of previously prepared lesson units which had been drawn on acetate visuals or transparencies and which had been mounted for use with an overhead projector. The acetate lesson units were accompanied by tape recorded sound and were designed to cover the basic instruction materials for French I. Two principle hypotheses were examined: 1) Students participating in an experimental course in French I, which made use of certain audio-visual and workbook materials and was taught in large part by specially trained student assistants, would demonstrate a degree of learning and achievement as great as that of a comparable group of students participating in the regular course in French I, but not using the audio-visual and workbook materials and taught in its entirety by the course instructor. 2) Students participating in such an experimental course would demonstrate a degree of satisfaction with the course and the instructor as great as that of a comparable group of students taking the regular course in French I.

The design of this study met high standards of scientific methodology.

The major findings of the study were as follows: 1) Data from both years of the study on a wide variety of measures indicated that students learned as well by the experimental method of instruction as they did by the conventional procedures. In those cases where significant differences did exist, these differences favored the members of the experimental group. 2) The new methods of instruction were received with considerable enthusiasm by the students. While there were few significant differences in the teacher rating scale, an analysis of individual items on this scale revealed that the members of the experimental groups rated the course and instructor more favorably than did the members of the control group on nine of twelve items on this scale. 3) The new study procedures offered major economies in the use of instructional time. Assuming an enrollment of sixty students in French I, French II, and French III, savings of instructional time are approximated at twelve hours per week for each level of French.

Further information about this study may be obtained by writing to Dr. Samuel Baskin, Director of Educational Research, Antioch College, Yellow Springs, Ohio.

NEED FOR MORE RESEARCH ON SUPERIOR TEACHERS

Mr. Byron Hansford, State Commissioner of Education, Colorado, appeared before a committee of the state legislature and asked for more research on means of identifying superior teachers. He said, "I would like to see some foundation give a grant of several million dollars to a university to determine this." He said further that results of such research could be as important as developing a vacine for polio.

FRIENDSHIP INTERRELATIONSHIPS

A study, accepted as a Ph. D. dissertation by the Department of Psychology, Princeton University, was conducted which tested the hypothesis that the more friendly students in a fraternity are, the more likely they would judge each other favorably on the eight dimensions: athletic ability, grade average, general intelligence, liking of good time, manners and experience, smoothness with girls, social background, and work for the fraternity. Two fraternity groups were studied; a four dimensional structure was found in each one. For one group the major friendship dimension was best indicated by athletic ability; for the other fraternity, the major friendship dimension was most highly related to intelligence and grades.

AN APPRAISAL OF THE Ed. D. PROGRAM AT WAYNE STATE UNIVERSITY

A study has been completed on the evaluation of the Doctor of Education program. Appraisal was made of instructional intent, recipient satisfaction, faculty evaluation, and dissertation quality.

THE UNIVERSITY OF LIBERIA

Dr. A. Gordon Nelson, Professor of Educational and Vocational Guidance at Cornell University, spent six weeks making a survey of the University of Liberia. A report has been prepared entitled "The University of Liberia: An Appraisal and Recommendations."

A Dembar Publicatio

The Journal of Educational Research Box 1605, Madison 3, Wisconsi

Devoted to the scientific study of education

JUST IN CASE YOU HAVEN'T HEARD:

You can make a substantial saving by entering or renewing your subscription to THE JOURNAL OF EDUCATIONAL RESEARCH now before our rates go up this fall.

And you'll certainly want to be with us in September, because The JER will be adding to its size and scope--with more pages and new features--the better to bring you rapidly and thoroughly the latest developments in basic investigations and advanced practices throughout education.

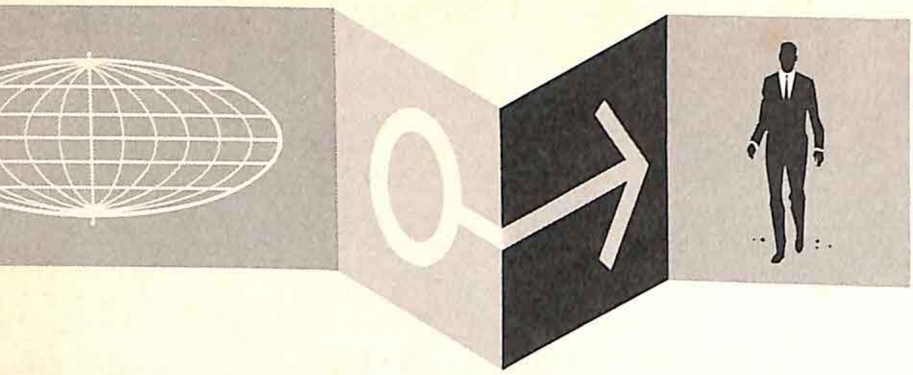
Yes, we'll have more full-length papers by distinguished scholars, a handy digest of "What's New in Educational Research," field news, book reviews, and other editorial departments designed to meet your needs for rapid orientation and thorough insight.

Sincerely yours,

A. S. Barr

A. S. BARR
Editor

PS: If you receive another announcement of our special offer, we'd appreciate your passing it on to a colleague.



The World, Th

At the growing edge of productive scholarship

THE WORLD OF EDUCATIONAL RESEARCH IS EXPLODING

It is almost commonplace to say so, but it is true. The growth of educational research is almost incomprehensible in its size and impact. In the past 10 years, more time, talent, and money have been invested in pushing outward the frontiers of educational knowledge than ever before in the history of man. In those same 10 years the findings of productive scholarship have had a profound influence on the scope and character of educational programs and methods. In the next 10 years we can confidently expect even more significant developments.

Devoted to the scientific study of education

THE JOURNAL OF EDUCATIONAL RESEARCH IS GROWING, TOO

In scope, in size, and in stature The Journal of Educational Research has sought constantly to improve its services to its chosen field. Over the past 15 years we have pioneered new formats that have made possible the more rapid and economical reproduction of more learned papers than had hitherto been thought possible. Now we are about to take another marked step forward — with even more pages devoted to selected research reports, new editorial features of added appeal and utility, and a revised typography that typifies the forward-looking role of The JER.

In Investigation and Interpretation

FOR 40 YEARS THE "MOST-QUOTED" JOURNAL

The Journal of Educational Research is the *first* journal in its field — the oldest, the most frequently cited in professional literature. Since 1920 The JER has been the recorder of progress in education. In that period the "frequency of use" of the major educational publications has twice been studied. Both times The JER ranked No. 1. In the past five years JER circulation has almost doubled. We think these a striking recognition of JER acceptance and coverage.

The scholarly journal with the common touch

FROM BASIC RESEARCH TO ENLIGHTENED APPLICATION

We can assure you that The Journal of Educational Research is not just another journal. It bridges the gap between fundamental investigation, applied scholarship, and practical application. The Journal reports findings, yes, but it also illustrates the ways that alert educators can study their own problems and devise improved practices. For both the technical work and the field worker The JER is a mine of ideas.

In the universe of education

A RUNNING RECORD OF LEADERS AND LEADERSHIPS

Traditionally the "big names" in education have added substance to The JER.

Through the years The Journal of Educational Research has presented early evidence on all the major break-throughs in education: individual differences, learning and learning problems, training techniques, tests and measurements, curriculum, counseling and guidance, methods of teaching in all subjects at all levels, supervision, investigations of reading, teacher education, evaluating teacher effectiveness, administration, the winning of public support, and other areas.

Journal, and You

By and for educators at all levels

WIDE OUTLOOK FOR A WIDE READERSHIP

The Journal of Educational Research belongs to no one country, no one institution, no one association. Its outlook is not circumscribed by geography or slant. The JER lives at the growing edge of productive scholarship in the universe of education, devoted to the discovery, documentation, and dissemination of the knowledge and insights by which new truths are found.

The readership of The JER testifies to its broad significance. Leading libraries are represented almost without exception. So are college and university administrators, professors of education, psychologists, elementary and secondary school specialists and teachers, vocational and agricultural education personnel, adult education leaders, research workers, educational associations, armed forces educational advisors, students, interested laymen.

Nearly 20% of The JER circulation goes overseas.

Discovery, documentation, dissemination

TOMORROW'S EDUCATION TODAY

Perhaps you used to be able to keep abreast of educational developments by reading annual abstracts or perennial digests. But no longer. The field of education is too swift and complex for that. Today's experiments become next week's concepts and next month's practices. Only by making up-to-the-minute research your "must" reading can you hope to stay in the forefront of your profession.

In every field there is a journal that stands for excellence, whose regular reading is a requisite. Now more than ever before, in the expanding universe of educational knowledge, *the* journal is The Journal of Educational Research.

JeR

Features:

Nine issues around the calendar: September, October, November, December-January, February, March, April, May-June, July-August.

Twenty percent more pages devoted to the latest research papers.

"What's New in Educational Research" . . . a monthly digest of late developments for the reader who needs rapid orientation and insight (and who doesn't?).

Thought-provoking editorials on emerging trends, problems, break-throughs.

News Summary . . . a page of highlights and sidelights on conferences, grants, appointments, projects.

Book reviews and article citations . . . providing a running over-view of the entire field of Education.

Critiques . . . "think pieces" by leading educators on the state of the discipline.

Look for these and other VITAL REPORTS

COMING IN SEPTEMBER: *How to Evaluate Teacher Performance.* A remarkable synthesis of the latest literature on this critical subject.

COMING IN FEBRUARY: *Why Johnny Can Read.* A collection of current papers on a topic of perennial significance.

COMING IN MAY: A special issue devoted to *Teacher Personnel Problems.*

PLUS periodic reports of new research on testing, learning, curriculum, administration, guidance, identifying the exceptional student, and many other areas.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 54 MAY, 1961 NUMBER 9

CONTENTS

Nonteaching Activities of Teachers in a Staff Utilization Study—Fred Guggenheim 325

Four Criteria of Principal Effectiveness—Walter R. Borg, Jack F. Burr, J. Arthur Silvester 332

The Relationships Between Categories of Behavior of Third Grade Teachers—George I. Brown 338

The Prediction of Teaching Performance—David L. Cole 345

Analysis of Attitude Inventory Data on Minnesota Vocational Agriculture Teachers—Gordon I. Swanson and Loyal W. Joos 349

A Follow-up Study of Teaching Candidates—J. C. Gowan, Carita Conner, and Phyllis Kennedy 353

Evaluation of Three Methods of Teaching Reading, Seventh Grade—Frederic R. Walker 356

INDEX

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

NONTEACHING ACTIVITIES OF TEACHERS IN A STAFF UTILIZATION STUDY

FRED GUGGENHEIM
University of the State of New York

IN THE FINAL analysis the true effectiveness of a school depends upon the teacher. Her well-being, training, and efficiency converge on the pupils to dictate a particular quality of teaching. This influence is exerted directly during the face-to-face contacts the teacher has with her pupils. Any evaluation of the efficiency and/or effectiveness of a school must consider these facts of prime importance.

However, there exist other factors associated with the teachers' role that have either a direct or indirect effect on the end product and which, by and large, have been neglected in educational research. These might be lumped together under the classification of nonteaching activities that are, although by varying degrees, related to the actual teaching process. One of these is the amount of time spent in teaching or, conversely, away from teaching. If teachers are, to a great extent, occupied throughout the day by nonteaching activities, this would have a direct bearing on the amount of actual teaching they do.

If the assumption of congruity of goals between schools and levels is accepted, then there should exist some relationships between schools and levels as to what teachers do or don't do. This assumption at the administrative level was found to be invalid.*

One of the questions raised in this study was whether any congruence existed on specific activities among teachers either between schools or between levels. All of the data collected were restricted to activities that were performed in addition to teaching. The rationale for this limitation was that a study of teaching activities should be coupled to a study of teaching methods and effectiveness. While nonteaching activities are related to effectiveness, the relationship is much more indirect and tenuous. Nonteaching activities also give a background or setting to the teaching function and offer another approach to building a theoretical model of school organization and function.

Many of the teacher activities investigated, such as lesson planning, are directly related to actual teaching activities. Others, like hall duty and school money collections are activities that are additional to teaching. In other words, certain activities are required precursors of teaching activities while others can or cannot be added without influencing the teaching process. This distinction is made in order to indicate that the nonteaching activities listed belong together only under a gross system of classification.

The data were classified and examined for high and low expenditure schools and between elementary and high schools.

Teachers were encouraged to give more than one response to a question. They were requested, however, to give them in order of importance. The tables used in this report contain first, second, and in some cases, third responses. The activities in each table were compiled from the teachers' responses.

Results

Regardless of school level, 64 percent of the teachers stated that they worked an average of eight hours per day in school. This is a surprisingly long time in view of the fact that most pupils attend school between 9:00 a.m. and 3:00 p.m., a total of six hours including lunch. This means that this sample of teachers spend 25 percent of their time in school engaging in activities that are other than face-to-face instruction. If free periods are taken into account, then the percent of time spent in actual teaching is shortened even further. It was found that high school teachers had an average of five and a half free periods per week and elementary school teachers had an average of three and a half free periods. Adding this to the 25 percent loss already reported would bring the total to above one-third of the school day being spent in nonteaching activities. The need for knowing how this time is accounted for becomes increasingly apparent.

---

* Kowitz, Gerald. Preliminary Study of Staff Utilization, The State Education Department, University of the State of New York, (unpublished research report) 1960.

The distribution of the number of hours per day teachers work and the number of free periods they have per week is presented in Tables I and II.

The number of hours teachers worked per day ranged from five hours to ten hours. There were no significant differences in the number of hours worked between elementary schools and high schools or between high expenditure and low expenditure schools.

No significant differences were found between high and low expenditure schools in the number of free periods teachers had. However, there was a significant difference in the number of free periods between elementary and high school teachers. Not only did high school teachers have more free periods, but the length of these periods was also greater. Free periods in the high schools were part of the teachers schedules, whereas in the elementary schools they existed only to the extent that special teachers took over the regular teachers classes for specialized instruction.

Each teacher was asked to name their five most important nonteaching activities and most time-consuming activities. Although they were all asked to name five, many teachers were only able to think of two or three nonteaching activities which they considered important. The activities and number of responses given are indicated in Tables III and IV.

Sixty-nine percent of the high school teachers listed sponsorship of pupil organizations and activities as their most important nonteaching activity. However, when the same group of teachers were asked what was their most time-consuming activity, only 49 percent of the responses were for sponsorship of pupil organizations, whereas 60 percent of the elementary school teachers responses indicated that the preparation of instructional material was the most time-consuming activity. The Spearman rank-order correlation was used to determine the extent of relationship between the activities teachers listed as most important and the activities listed as most time-consuming. A correlation coefficient of .81 was obtained which was significant at the .01 level. The distributions in Tables III and IV are of first, second, and third responses. Although the correlation between the most important activities and the most time-consuming activities is high, important individual discrepancies still exist. An example of these discrepancies can be seen in Table V from a tabulation of first responses to two of the highest ranked items.

The importance of an activity in an educational organization does not necessarily determine the amount of time needed in its performance. The nature of the activity would be most crucial in dictating the time needed. In the example cited, preparation of instructional material is a constant, ongoing activity, whereas sponsorship of pupil organizations usually occurs during the school day with definitely established time limits.

The same group of teachers were asked to name three nonteaching activities which they felt needed more time and three which they felt should be dropped. The activities which teachers named as deserving more time are presented in Table VI.

In all classifications teachers responses were internally consistent and also consistent with those activities which they reported as being most important. Preparing instructional materials and individual work with pupils accounted for 55 percent of the responses. A conflict in educational activities becomes apparent when these responses are compared to the responses in Tables III and IV. In the high schools, sponsorship of pupil organizations was reported as being the most important nonteaching activity and the second most time-consuming activity. Individual work with pupils was ranked as fifth in importance and seventh in time consumption. However, it was ranked first among activities which needed more time.

The meaning of this disparity becomes clearer when the data are examined against the backdrop of the school organization. Nonteaching activities, such as pupil organizations, study hall, monitorial duties, etc., are established through administrative policy. Preparation of instructional materials and individual work with pupils are executed on an individual basis. They have little bearing on school organization or class scheduling. It becomes questionable as to how important teachers really feel sponsorship of pupil organizations is when this activity is compared to those activities which they feel deserve more time. There were only six first or second responses listing this activity as deserving more time.

There are several explanations that could account for some of the apparent discrepancies existing among activities considered to be the most important, the most time-consuming, and deserving of more time. For one thing, each question might have been answered against different frames of reference. For example, sponsorship of pupil organizations might have been reported as the most important nonteaching activity because the administration considers it the most important or because it occupies most of the teachers nonteaching time. It might be reported as the most important nonteaching activity, yet not be even mentioned as deserving more time. This might be because it receives sufficient time or because of other, more subtle reasons. Before definitive statements can be made or conclusions reached concerning nonteaching activities, especially what teachers believe about them, further study in depth must be made. Some of the conflict in the data is more apparent than real, probably being due to the varying degrees of amalgamation in the responses of facts with attitudes.

TABLE I

HOURS WORKED IN SCHOOL PER DAY

	5		6		7		8		9		10		11		No Data	
	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
High Schools	1		2		7	1	9	18	3	4	1					
Elementary			1	2	2	2	9	7	1	5		1				
Combined	1	0	3	2	9	3	18	25	4	9	1	1				

TABLE II

FREE TIME IN PERIODS PER WEEK

	None		1 - 2		3 - 4		5 - 6		7 - 8		9 - 10		11+		No Data	
	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
High Schools	2	1	2		3	1	8	16	4	1	5	3	2			
Elementary		2	3	5	8	2	4	3		1	1					
Combined	2	3	2	1	7	3	12	19	4	2	6	3	2			

TABLE III

MOST IMPORTANT NONTEACHING ACTIVITY

Activity	High School		Elementary		Combined	
	High	Low	High	Low	High	Low
Preparing instruction materials	5	8	8	9	13	17
Individual work with pupils	5	3	3	6	8	9
Conferences		2	1	2	1	4
Correcting homework and class tests	3	10	4	2	7	12
Personal study and planning	3	2	1	4	4	6
Official class reports		5	2	1	2	6
Other official reports	2				2	
Counseling	2	2	1		3	2
Monitorial duty	1	5	2	3	3	8
School collections		1	1		1	1
Administrative assignments	7	2	5	4	12	6
Faculty and department meetings	2	1	4	2	6	3
Study hall	1	4			1	4
Coaching athletics	2	1		2	2	3
Sponsorship of pupil organization and activities	14	13	1	2	15	15
Field trips			1		1	
Teachers association activities	2	2	4		6	2
Homeroom	2	1	1		3	1
None	2	1	1		3	1
Discipline		1				1

TABLE IV

MOST TIME-CONSUMING ACTIVITY

Activity	High School		Elementary		Combined	
	High	Low	High	Low	High	Low
Preparing instruction materials, lesson plans	12	15	10	6	22	21
Individual work with pupils	2	3	2	1	4	4
Conferences			1			1
Correcting homework and class tests	10	15	8	8	18	23
Personal study and planning	3	3	1	1	4	4
Official class reports	4	4	4	8	8	12
Other official reports						
Counseling	1	2			1	2
Monitorial duty		1	2	4	2	5
School collections	1		5	1	6	1
Administrative assignments	5	1	2	2	7	3
Faculty and departmental meetings	2	1	3	2	5	3
Study hall	1	2			1	2
Coaching athletics	1			1	1	1
Sponsorship of pupil organizations	10	9		2	10	11
Homeroom	2	1			2	1
Obtaining supplies	1				1	
Audio-visual	1				1	
Discipline		1				1
Teachers Association		1	1		1	1

TABLE V

EXAMPLES OF DISCREPANCIES IN HIGH SCHOOL TEACHERS FIRST RESPONSES

	Most important	Most time-consuming
Preparing instructional material	7	12
Sponsorship of pupil organizations	15	5

TABLE VI

ACTIVITIES WHICH DESERVE MORE TIME

Activity	High School		Elementary		Combined	
	High	Low	High	Low	High	Low
Preparing instruction materials	7	7	9	7	16	14
Individual work with pupils	9	9	8	7	17	16
Conferences		4	3	2	3	6
Correcting homework and class tests	1	3	1	1	2	4
Personal study and planning	2	6	1	1	3	7
Official class reports	1		1		2	
Other official reports						
Counseling	1		1		2	
Monitorial duty						
School collections						
Administrative duties	3	2	1		4	2
Faculty and departmental meetings	1				1	
Study hall						
Coaching Athletics						
Sponsorship of pupil organizations	4	2			4	2
Visitation	1				1	
Field trips	1		1		2	
Teachers Association	1	1			1	1
None	1	3	2	1	3	4

Conclusions and Summary

As a result of this study, certain questions and hypotheses are raised concerning the relationship between administrative goals and teachers activities. Kowitz*, in a study of congruence in administrative policy, using the same school population as in this study, found little or no agreement among administrators on educational policy. Contrary to his findings as to the lack of agreement among administrators, the teachers in the same schools exhibited a marked agreement among themselves in what they reported. The congruity was found, by and large, between levels and between schools. A lack of congruity was found among activities.

The results of this study and those of Kowitz indicate a marked incongruity existing in school functioning. One would expect, at least to a noticeable degree, that if there was dissonance present at the administrative level, an equal or greater amount at the staff level. This would be because of the downward flow of educational policy. That this was not found is all the more remarkable. Several conclusions lend themselves to these results. However, because of the relatively few schools and personnel involved in the study, the conclusions are used to raise certain questions that need further study.

At first glance the results indicate that teachers engage in activities that are fairly uniform regardless of administrative policy. This might be due to similarity of training, meeting of minimum legal requirements, or because there is no relationship between the administrative policy studied and the activities that teachers were queried on. It should also be kept in mind that the data collected was from what the staff said they did and not what they were observed doing.

The congruity found at the teaching level can, at least in part, be explained by an inherent tendency of the units themselves. This can best be illustrated by an analogy to a building structure composed of multiple components. It can readily be seen how given units can be organized into a variety of patterns to yield similar or dissimilar, depending on the architect's inclinations, structures. However, among the components themselves there exists a tendency to maintain individual integrity. By defining the teacher and her pupils as the components of the school structure, the aforementioned logic can be applied. Indeed, the teacher and pupils, taken together, are the single unit that has remained historically constant.

The question of incongruity among activities was discussed in some detail in the body of this report. It was found that different, but logically related, questions about the same activities resulted in apparent conflicts in responses. The hypothesis was raised that the conflict might be a result of obtaining responses that were mixtures of varying degrees of facts and attitudes.

The results of this study, individually and compared with other sections of the staff utilization study, indicate that research on teachers' activities presents several multidimensinnal and complex problems. Any research design in the future should take congizance of the interrelationships between the reporting of facts and the influence of attitudes, values, and expectations upon these facts. This is illustrated by a comparison of two questions—what do you do, as compared to, what is the most important thing you do?

---

* Ibid.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

FOUR CRITERIA OF PRINCIPAL EFFECTIVENESS

WALTER R. BORG, JACK F. BURR, J. ARTHUR SILVESTER
Utah State University

PERHAPS ONE OF the most important problems in educational administration is the lack of satisfactory criteria of administrator effectiveness. This study is concerned with four functional criteria of principal effectiveness.

Review of the Literature

The need for research in the field of school administrator criteria has been emphasized by a number of authorities (7, 10, 11, 12, 16). In spite of the obvious need for research in these fields, however, the great bulk of work published to date contributes little to the establishment of sound criteria. Most of this work consists of accumulations of opinion in one form or another, usually involving the preparation of lists of qualities needed by educational administrators (1, 2, 4, 6, 7, 3, 9, 12, 13, 14, 15). Some of these lists arise from surveys based upon national samplings, some upon the opinion of a single author, some upon commitee reports, while others are based upon opinions of school teachers in a given high school or school district. These studies make a real contribution to our thinking concerning the nature of the administrator's responsibilities and are certainly valuable as far as they go. Their weakness, however, is that they provide no sound, measurable yardsticks against which we can weigh new systems of administrator evaluation, selection, or prediction.

Method

The present study made extensive use of the many lists of school administrator characteristics available in the literature. These were coded and combined in order to evolve a composite picture which it was felt would be more representative of overall professional opinion concerning administrator characteristics than would be any single study conducted in this field. After some 35 of these studies were coded and combined into a composite characteristics list, this list was used as a basis for developing rating and observational instruments to be used in obtaining the four functional criterion measures compared in this study.

Criterion data were gathered on 27 elementary principals and 1 high school principal, from three districts.

The four functional criteria employed in this study consisted of a superintendent ranking of all principals under the supervision of the superintendent, a teacher rating in which all teachers in a principal's school were called upon to render an anonymous evaluation of their principal, a rating based on the evaluation of the principal's performance in a faculty meeting as scored by two independent observers, and a self-rating rendered by each principal. As it was felt that the position of the rater would have an influence on those characteristics that he could intelligently evaluate, the ratings were not identical. Most of the characteristics chosen, however, were observable by all raters and were thus found on all scales.

Table I gives a listing of all characteristics rated with check marks indicating which characteristics were included in each of the criterion measures. It will be noted that some areas were combined in the faculty meeting observation in order to simplify the work of the observers. The area of "Teacher Attitude" was added to the faculty meeting rating. Observers attempted to estimate the teacher's attitude toward the principal from his behavior in the faculty meeting. As this variable was not related to the composite teacher rating of the principal, it appears that observers were not successful in appraising teacher attitude.

Findings

It was possible to compute reliabilities on two of the four criterion measures. In the case of teacher ratings, the teachers in a given school were randomly divided into two equal groups and the ratings of one group correlated against the ratings of the other group, thereby providing a split-half reliability estimate of teacher ratings. Intercorrelations, using the rank difference technique, were computed and the results are shown in Table II. The reliability of the overall teacher ratings was found to be .65 while the reliabilities of specific rating areas may be seen to fluctuate between .04 and .65. It is

TABLE I

VARIABLES RATED ON THE FOUR CRITERION MEASURES

Variable	Superintendent Rank	Teacher Rating	Faculty Meeting Rating	Self-Rating
1. Personality	√	√	√	√
2. Administrative Ability	√	√	√	√
3. General Knowledge	√			
4. Professional Knowledge	√			
5. Cooperativeness	√	√	√ **	√
6. Tact	√	√	√ **	√
7. Stimulates Co-workers	√	√	√ **	√
8. Socially Active	√	√	√ **	√
9. Good Judgment	√	√	√ *	√
10. Originality	√	√	√ *	√
11. Communicative Skills	√	√	√	√
12. Forcefulness	√	√		√
13. Physical Characteristics			√	
14. Teachers' Attitude			√	

* Combined under Intellectual Ability

** Combined under Human Relations Skills

TABLE II

RELIABILITIES OF TEACHER RATINGS

Variable	Rho
Cooperativeness	-.04
Administrative Ability	.42
Personality	.65
Stimulates Co-workers	.62
Communicative Skills	.46
Socially Active	.40
Good Judgment	.25
Originality	.29
Tact	.47
Forcefulness	.33
Total	.65

TABLE III

INTERRATER RELIABILITY OF THE FACULTY MEETING OBSERVATION

Variable	Rho
Human Relations	.56
Communicative Skills	.63
Teachers' Attitude	.95
Personality	.78
Physical Characteristics	.72
Intellectual Ability	.56
Administrative Ability	.71
Total	.78

interesting to note that such variables as personality and ability to stimulate co-workers can be rated with reasonably good reliability by the teachers while such variables as cooperativeness, good judgment, and originality show very little reliability.

It was also possible to compute interrater reliabilities for the faculty meeting observations as two independent raters were employed in this activity. These raters were trained in the use of the faculty meeting observation sheet prior to the start of the study which probably explains the somewhat better reliability than was obtained with the teacher ratings. The reliability of the total faculty meeting observation rating was .78, with the reliabilities of specific subareas of the rating scale ranging from .56 to .95 (See Table III). Ratings of human relations skills and intellectual ability were the least reliable in the faculty meeting observation scale while ratings on teacher attitude, personality, physical appearance, and professional skill were of satisfactory reliability.

Correlations were computed between the total scores of the 28 principals on the four criteria as well as on the specific subareas which occurred on more than one of the criterion measures. Because of the small number of cases involved, the rank difference correlation was employed in these comparisons.

The comparison between the faculty observation ratings and teacher ratings may be found in Table IV. It may be seen that the correlations between the two total ratings is only .45 with correlations ranging from -.08 to .51 for specific subareas that were included on both of these scales. With the instruments employed in this study it is apparent that the agreement between these two criterion measures is low. Correction for attentuation, which provides an estimate of the maximum relationship that might be obtained with reliable measures, raises the correlation between total scores, administrative ability, intelligence, and personality to .65 or above. It was thought that the teacher's attitude as observed in the faculty meeting would be related to the total teacher rating of the principal but this correlation was not significant. The total teacher rating did correlate significantly, however, with physical characteristics of the principal, such as energy, attractive appearance, and poise.

Comparison between faculty meeting observations and superintendent ratings yielded rank difference correlations ranging from .34 for teacher attitude to .56 for cummunicative skills (see Table IV). The correlation between the overall faculty observation rating and overall superintendent rating, however, was only .40 indicating that relatively little common variance is being measured by these two criterion measures. As no estimates for superintendent rating reliability could be made, it is not possible to make corrections for attentuation.

A comparison of the rating made by the principals' teachers with ratings made by the principals' superintendent seems particularly interesting as both of these evaluations are important to the functioning of the principal in the school system. There is no reason to expect a high level of agreement to be present between superintendent and teacher evaluations as the superintendent certainly sees the principal in a markedly different role than the teacher. Thus, it is not surprising that the correlation between overall teacher rating and overall superintendent rating is low (see Table IV).

Some of the correlations between specific characteristics as rated by the teachers and the superintendents are interesting. It may be seen that there is reasonably good agreement (considering the reliabilities of the ratings) between the teachers and superintendents in the areas of communicative skills (r = .65), level of social activity (r = .54), and administrative ability (r = .55). The correlation of -.02 between the teacher rating and superintendent rating of the principal's judgment certainly suggests that the two criterion groups view the principal quite differently in this area. It seems likely that the types of principal judgments perceived by the superintendent are considerably different from those judgments observed by the teacher in the principal's day-to-day contacts with them at school.

Relationships between the fourth criterion, the principal's self-evaluation and the other three criteria were so low that detailed treatment is not appropriate. Correlations between the overall self-perception score and overall ratings in the other three criteria range from -.02 to .19, suggesting that no significant relationship exists between the principal's self-perception and the way he is perceived by superintendents, teachers, and faculty meeting observers. The results of the self-ratings seem to suggest that as a group the principals in this research displayed relatively little insight concerning their professional capabilities, etc. It appears doubtful that the self-rating can be developed into a satisfactory criterion.

Summary and Conclusions

Based upon a composite of desirable characteristics of principals as listed in 35 studies, a series of four criterion measures were developed. These measures consisted of a superintendent ranking, a teacher rating, a principal rating based on faculty meeting observations, and a self rating rendered by the principals involved in the study. Comparisons of the scores earned by 28 principals on these four criterion measures indicate relatively little common variance between the various criterion measures. Reliabilities of the measures are, in many cases, low and the development of more reliable measures would raise the relationships among the various criteria to a considerable degree. In spite of the generally low relationships between the criteria, however, the criterion measures agreed to a satisfactory degree in certain specific rating areas.

This study certainly indicates that studies in-

TABLE IV

CORRELATIONS BETWEEN THREE FUNCTIONAL CRITERIA

Variable	Faculty Meeting Rating vs. Teacher Rating	Faculty Meeting Rating vs. Superintendent Rank	Teacher Rating vs. Superintendent Rank
Administrative Ability	.51**	.40*	.55**
Intellectual Ability^{x}	.33	.22	
Originality			.13
Good Judgment			-.02
Human Relations Skills^{xx}	.03	.27	
Cooperativeness			.30
Tact			.38*
Socially Active			.54*
Stimulates Co-workers			.43*
Personality	.42*	.37*	.28
Communicative Skills	.32	.56**	.69**
Teacher Attitude***	-.08	-.34	
Physical Characteristics	.45*	.43*	
Forcefulness			.26
Total	.45*	.40*	.28

* Significant at the 5% level.

** Significant at the 1% level.

*** These scores on faculty meeting observation were correlated with total teacher rating.

x The Intellectual Ability score was arrived at by combining two subscores: Originality and Good Judgment on Teacher Rating and Superintendent Rating.

xx The Human Relations score was arrived at by combining four subscores: Cooperativeness, Tact, Socially Active, and Stimulates Co-workers on Teacher Rating and Superintendent Rating.

volving different criterion measures of principal success cannot be compared with any degree of confidence. As far as the relative merits of the criterion measures themselves, it is impossible to determine this from the results of this research. It is possible, excepting the self-ratings which seemed to be related to nothing else, that the other criterion measures are all reasonably valid and yield different results merely because the principals are being measured in different roles. Certainly the responsibilities of the principal involve a number of roles in which he must behave and his behavior in one role may not necessarily be closely related to his effectiveness in his other roles. A composite criterion measure might be developed from these three measures which would provide a reasonably valid overall criterion of principal effectiveness.

Although criterion measures similar to those employed in this study have been used successfully in other areas of job evaluation, it appears that the roles and behavior requirements for school administrators are so varied as to render simple rating criteria of the sort employed by this study basically inadequate. In all likelihood, what is needed in the development of criteria of administrator effectiveness is a bold behavior-oriented approach which would include measures of human relations skills, effectiveness in dealing with administrative problems, and group interaction effectiveness using role playing, leaderless group discussion, and similar techniques taken from the fields of psychology and sociology. Certainly, we have ample evidence from education and other fields to indicate that such unimaginative criteria as college grades in educational administration courses would be almost entirely useless for predicting successful performance in the field (9).

REFERENCES

1. Anderson, Walter A. and Hunnicut, C. W., "Leadership Qualities and Competencies," School Executive, LXIX (July, 1950), pp.48-49.
2. Campbell, R. F., "Research and the Selection and Preparation of School Administrators," Educational Research Bulletin, XXXV, No. 2 (February, 1956), pp. 29-33.
3. Campbell, R. F. and Gregg, R. F. (Ed.), Administrative Behavior in Education. New York: Harper Brothers, 1957, 547 pp.
4. Cohler, M. J., "Faculty Helps Select the Assistant Principal," American School Board Journal, CXVIII (February, 1949), pp. 33-34.
5. Coladarci, Arthur P., "Administrative Success Criteria," Phi Delta Kappan, XXXVII (April, 1956), pp. 283-285.
6. Enochs, James B., "Elementary School Administrators Evaluate Themselves," Elementary School Journal, LI (September, 1950), pp. 15-21.
7. Forester, J. J., "Selecting and Training Elementary School Principals," National Elementary Principal, XXXIV, No. 2 (October, 1954), pp. 32-35.
8. Grobman, H. and Hines, V. A., "What Makes a Good Principal?" National Association of Secondary School Principals Bulletin, XL (November, 1956), pp. 5-16.
9. Kearney, M. E., "What Principals Must Be Made of," Nations Schools, LIX, (June, 1957), pp. 57-60.
10. Moore, H. A., Jr., Studies in School Administration: A Report on the Conference of Professors of Educational Administration. Washington, D. C.: American Association of School Administrators, 1957, 202 pp.
11. Ramseyer, J. A., "Administration Program Development," Phi Delta Kappan, XXXVII (April, 1956), pp. 299-303.
12. Ramseyer, J. A. and Staff, Developing Guidelines for Educational Leadership, (Fifth Annual Report) School-Community Development Study. Columbus, Ohio: Ohio State University, 1955-1956, 91 pp.
13. Roncans, R., "Eight Planks of Elementary Principalship," School Executive, LXXVI (September, 1957), pp. 72-73.
14. Templeton, A., "Yonkers System of Selecting Principals," School Executive, LXXI (June, 1952), p. 61.
15. Wetzler, W., "The Marks of a Successful School Administrator," American School Board Journal, CXXXI (November, 1955), pp. 42-43.
16. Worth, C., "The Selection of a High School Principal," American School Board Journal, CIII (July, 1941), pp. 49, 68.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

THE RELATIONSHIPS BETWEEN CATEGORIES OF BEHAVIOR OF THIRD GRADE TEACHERS

GEORGE I. BROWN
University of Delaware
Newark, Delaware

MANY ASSUMPTIONS have been made about teacher behavior. For example, it is often assumed that a teacher who is given to using predominantly directive statements in the classroom rather than a more pupil-centered problem structuring approach in her teaching is usually more "strict" and would thus use more reproof than praise with her pupils. The converse is, of course, also assumed.

The purpose of this study is to present data concerned with the interrelationships of various categories of teacher behavior in order to suggest whether certain "common sense" assumptions such as those stated above have any empirical basis.

The measurements made in the study were those of categories of teacher behavior essentially verbal in nature. These verbal categories could be conceived as approximations of the teacher's total classroom behavior or may be considered only in terms of the category descriptions; similar implications concerning the relationships between categories would probably be drawn in either case.

Instrument

A technique developed by John Withall (3, 4) to measure classroom climate was used to measure and categorize teachers' verbal behavior. Withall separated all teacher verbal response into seven categories. His definitions of these categories were slightly refined in this study. Supplementary comments are included along with Withall's original definitions as follows:

Category 1. "Learner-supportive statements or questions.

These are teacher statements or questions that express agreement with the ideas, actions or opinions of the learner, or that commend or reassure the learner. Agreement is frequently expressed by a mono-syllabic response such as 'Yes,' 'Right,' 'Uhhuh,' and the like. Commendation or reassurance may be stated in terms of:

a: class accepted criteria or goals
or
b: the private goals and subjective criteria of the teacher.

The dominant intent of these statements or questions is to praise, encourage or bolster the learner." (4:96)

Sometimes teacher responses in this category combine two or even all three of the above purposes in the intent of a response.

For example:

1. Why that's fine, Tommy. Keep up the good work!
2. Yes, now you're getting it. You certainly got that one quickly.
3. Jane, you are writing better every day.

Responses in this category are relatively easy to identify if the observer holds fast to intent as the criteria for evaluation. The words "That's fine," can range in meaning from a sarcastic "What a stupid thing to do," to paens of praise meaning, "What a wonderful, superb piece of work you have done." There are times when the emotional tone alone is not enough and the observer must depend on the contextual situation for clues as to intent. The categorization of this and every category depends on the total classroom situation or "field" at the time of the response.

Category 2. "Acceptant or clarifying statements or questions.

These are teacher statements or questions which either:

a: accept, that is, evidence considerable understanding by the teacher,
or
b: clarify, that is, restate clearly and succinctly in the teacher's words the ideational or the feeling content of the learner's statement. The dominant intent of these teacher-responses is to help the learner to gain insight into his problem, that is, define his 'real' problem and its solution in more operational terms." (4:96)

Some typical examples follow:

1. Pupil: When I put the three on top of the

two instead of the two on top of the three it wouldn't change the sum.
Teacher: You mean the order of the numbers doesn't make any difference when you add.

2. Pupil: If we could only take a trip to South America and see how the children there live then we'd really know.
Teacher: The best way to learn about South American children probably would be to actually live with them.

3. Pupil: Why do we have to practice handwriting every day? It gets so, so, - boring.
Teacher: Sometimes you can't see any reason why we have to practice our handwriting every day, and you wish we didn't have to.

It may be seen from the examples given that acceptance and clarification sometimes occur concurrently, as in examples 1 and 2. Probably statements made for the purpose of clarification must, of necessity, give evidence of acceptance, if acceptance is thought of as giving "evidence (of) considerable understanding." It is possible, however, to have pure acceptance without an accompanying intent of clarification as in example 3.

Category 3. "Problem-structuring statements or questions.

Problem-structuring responses by the teacher offer facts or ideas or opinions to the learner about

a: phenomena,

b: procedures,

in a non-threatening and objective manner. These responses contain NO element of advising or recommending the adoption of certain ideas or procedures. Problem-structuring responses are frequently posed as questions which seek further information from the learner about the problem confronting him; or they may be statements which offer information to the learner about his problem. The learner is free to accept or reject in part or in entirety the facts or opinions that are presented to him. Problem-structuring responses may be questions which the teacher asks (1) to further increase his own understanding of what the learner has said, or (2) to increase the precision of this learner's statement of the problem. Problem-structuring responses are problem-centered rather than either teacher or learner-centered; nevertheless, they do tend to sustain the learner by facilitating his problem-solving activities. The dominant intent is to elucidate the problem and therefore facilitate the learner's problem solving activities." (4:96-97)

The identification of responses in this category is made easier if the observer remembers that the teacher's responses are directed at structuring the learner's problem. This is true without regard to how or why the problem became the learner's. It is conceivable that immediately preceding such a response, the teacher could have insisted a pupil solve an example in subtraction as illustrated below:

1. Teacher: John go to the board and do the first example.

2. Pupil (going to the board): I don't know how to begin.

3. Teacher: Do you remember what we did yesterday?

4. Pupil: Yes.

5. Teacher: Is there anything about today's example that is similar?

6. Pupil: You mean change the tens to ones?

7. Teacher: What would happen if you did?

Here, teacher statement 1 imposes a problem upon the learner. But teacher responses 3, 5, and 7 would in any case fall in Category 3 as problem-structuring questions.

It is important for the observer to keep in mind that responses falling into Category 3 focus primarily on the problem of the learner, not a problem of the teacher. The property of being concerned with a learner's problem is essential to the definition of this category.

A large percentage of teacher responses will probably fall in this category.

Category 4. "Neutral statements evidencing no supportive intent.

These statements are neither teacher-sustaining, nor learner-sustaining nor problem-centered. They constitute a very small percentage of the total teacher-responses. These responses include statements in which the teacher: (1) questions himself aloud; (2) repeats verbatim a statement that the learner just made; (3) uses a polite formality, et cetera. Statements having to do with administrative procedure--the room in which the class will meet, the hour at which a conference will occur, especially after concensus has been achieved, fall into this category." (4:97)

Some examples are:

1. At eleven o'clock, we will go to assembly.

2. Who knows where the yardstick is?

3. My, it's certainly cold out today.

Relatively few teacher responses fall in this category. Withall's definition adequately covers the content and limits of this category.

Category 5. "Directive statements or questions.

These are teacher-statements or questions which advise the learner regarding a course of action or his future behavior and which narrowly limit his choice or offer no choice. These statements recommend to the learner the facts or procedures that the teacher proffers him. These statements or questions convey the impression to the learner that the teacher expects and hopes that he will follow her prompting and that she will approve if he does. The dominant intent of these responses is to have the learner take up the teacher's point of view and pursue a course of action that she advocates. "(4:97)

Again, as in Category 3, a large percentage of teacher responses will probably fall in this category. Some examples are given below:

1. Jean, take your seat and open your book to page 10.

2. Be sure to tell the reasons why you liked the book.

3. Read quietly and when you have finished, hold up your hand.

4. Tell me why Jimmy didn't like the circus.

5. Are you going to finish your work?

Examples 4 and 5 might easily fall in other categories if they are spoken with different teacher intent. But if they are teacher-centered in intent they belong in Category 5. This illustrates the need to hear the emotional tone of a response as well as observe the context in which it is made.

To repeat, the observer will be assisted in identifying those responses belonging in Category 5 by remembering that it is essential that these responses display the primary intent of being teacher-centered as opposed to the responses of Category 3 which center on the pupils' problems.

Category 6. "Reproving, disapproving or disparaging statements or questions.

By means of these statements a teacher may express complete or partial disapproval of the ideas, behavior and, to him, personality weaknesses of the learner. The teacher's internalized societal values largely enter into these responses. By means of these statements some teachers believe they are fulfilling their responsibility of inculcating in young people society's standards of acceptable and desirable behavior and achievenent.

The dominant intent of these statements is:

a: to represent to the learner societal values as the teacher sees them;
b: to admonish the learner for unacceptable behavior and to deter him from repeating it in the future;
c: to impress on the learner the fact that he has not met the criteria for successful achievement which the teacher accepts. " (4:97)

To these might be added those responses whose purpose it is:

d: to strike out at those learners whose behavior has been perceived by the teacher as a threat to her;
e: to get relief from the teacher's own "bottled up" pressure by using a learner as an object to "let off steam" at, etc.

It might be in order to clarify Withall's comments on the teacher's internalized societal values. It is when the pupil's behavior--past, present or potential--is perceived by the teacher as being in conflict with those values that an appropriate Category 6 response might be made

Examples of Category 6 responses:

1. I don't want to see you do that again!

2. Don't be so silly!

3. I've never seen a class act the way this one does.

4. Sit down and be quiet!

5. Now, you know better than that.

Responses in this category should not be difficult to identify as to dominant intent. It can be seen that they are the antithesis of Category 1, learner-supportive responses.

Category 7. "Teacher-supportive statements or questions.

These are statements or questions in which the teacher refers to himself and expresses a defensive attitude, or refers to his present or past interests, activities or possessions with the purpose of reassuring himself and of confirming his position or his ideas in the eyes of those around him. The dominant intent of these statements is to assert, to defend, or to justify the teacher. Statements in which the teacher perseverates on an idea, a belief or a suggestion would fall in this category. By 'perseveration' is meant a persisting in, a reiteration of, and rigid advocacy of an idea or opinion by the teacher despite additional data being presented to her which calls for a reexamination of the original idea or opinion. " (4:98)

TABLE I

TEACHERS' MEAN SCORES* OF CATEGORIZED BEHAVIOR

Teacher	Cat. 1	Cat. 2	Cat. 3	Cat. 5	Cat. 6	Cat. 7	Cats. 1+2	Cats. 6+7	Cats. (1+2)-(6+7)+ 30
A	19.5	1.0	58.0	14.0	6.0	0.5	20.5	6.5	44.0
B	18.5	3.0	54.5	17.0	3.5	0	21.5	3.5	48.0
C	29.5	0	47.0	20.5	2.5	0	29.5	2.5	57.0
D	9.0	3.5	59.5	20.0	9.0	0	12.5	9.0	33.5
E	13.0	4.0	55.0	20.0	9.0	0	17.0	9.0	38.0
F	20.0	2.0	47.0	21.0	9.0	1.0	22.0	10.0	42.0
G	17.0	1.0	49.0	22.5	6.5	0.5	18.5	7.0	41.5
H	8.5	6.5	49.5	20.5	13.5	0.5	15.0	14.0	31.0
I	11.5	1.5	50.5	28.5	7.0	0.5	13.0	7.5	36.5
J	12.5	0	48.0	23.5	14.5	2.0	12.5	16.5	26.0
K	16.0	1.0	36.0	32.0	13.0	1.0	17.0	14.0	33.0
L	17.0	0	32.0	26.0	20.0	1.0	17.0	21.0	26.0
M	8.0	0.5	35.0	35.0	15.0	5.5	8.5	20.5	18.0
N	10.0	0	22.0	36.0	26.0	5.0	10.0	31.0	9.0
O	4.5	1.0	27.0	44.0	19.0	1.0	5.5	20.0	15.5

* Percentage of total verbal responses falling into a category.

TABLE II

BEHAVIOR CATEGORY INTERCORRELATIONS (PRODUCT-MOMENT)

	1. Cat. 1	2. Cat. 2	3. Cats. 1+2	4. Cat. 3	5. Cat. 5	6. Cat. 6	7. Cat. 7	8. Cats. 6+7	9. Cats. (1+2)-(6+7)
1. Cat. 1 (Pupil-Supportive)		-.31	.96	.31	-.58	-.60	-.42	-.60	.81
2. Cat. 2 (Acceptant)	-.31		-.04	.52	-.37	-.24	-.35	-.30	.16
3. Cats. 1+2	.96	-.04		.48	-.73	-.71	-.54	-.72	.90
4. Cat. 3 (Problem Structuring	.31	.52	.48		-.86	-.83	-.61	-.85	.74
5. Cat. 5 (Directive)	-.58	-.37	-.73	-.86		.70	.55	.72	-.77
6. Cat. 6 (Reproof)	-.60	-.24	-.71	-.83	.70		.62	.98	-.93
7. Cat. 7 (Teacher-Supportive)	-.42	-.35	-.54	-.61	.55	.62		.74	-.71
8. Cats. 6+7	-.60	-.30	-.72	-.85	.72	.98	.74		-.95
9. Cats. (1+2)-(6+7)	.81	.16	.90	.74	-.77	-.93	-.71	-.95	

As in Category 2, many of the responses which might belong in this category might be interpreted by observers as a Category 6 response because they are usually also intended to reprove, etc. However, the unique characteristic they possess is that of the teacher feeling the need to defend something she had done or said, or even thought she had said or done previously. The following are examples of this category:

1. If you had listened to me you would have done it right in the first place!
2. I don't care what Mrs. Johnston did, in this room you . . .!
3. Well, I didn't understand what you said. If you'd speak more clearly in the first place we'd all understand you better!
4. Russians are stupid. That's all there is to it and I don't want to hear any more about it!

X. That's enough.* 5. I don't want to hear any more arguments. We'll do it my way.

Not many responses in this category seem to to be used by most teachers.

The next problem facing the observer who decides to use Withall's technique is one of criteria for determining what one complete and distinct response is.

A response may be a single word or a combination of words, often more than one sentence. In this study, a response must express or imply a single, unified, complete and distinct intent within the total classroom context in its utterance. Each response in context can stand alone on its unique meaning.

Sample and Procedure

The sample consisted of fifteen third grade teachers. Reliability between two observers was established during the observation of six of these teachers. (There were an insufficient number of Category 4 statements to justify the inclusion of this category in the study.) The correlations between the observers' measurements were as follows:

Category 1 statements	.74	Category 5 statements	.90
Category 2 statements	.82	Category 6 statements	.92
Category 3 statements	.93	Category 7 statements	.78

After observer reliability was established the remaining observations were made individually by the observers. The measurements used in the intercorrelations in Table 2 were confined to those made during the teaching of reading. (This was due to the nature of the parent study (1).) Each teacher was visited at least twice at weekly intervals for an observation period covering about one hundred teacher responses.

Table 1 presents the mean percentages of categorized statements for each teacher. It will be noted that (a) Category 1 and Category 2 and (b) Category 6 and Category 7 are combined. In (a) this was done to form a more comprehensive measurement of praise. Both categories here could be interpreted as having praise as the teacher's intent. Similarly, in (b), the total percentage of responses falling into both Category 6 and Category 7 could be defined as having reproof as part of the intent of the teacher.

The difference between these two combinations of categories is also included in Table I. This reveals how much more praise than reproof, or vice versa, was used by each teacher. (The constant 30 is added to eliminate negative values in the table.)

The reader will note the range and relatively even distribution of the scores in the columns. The teachers are listed in order of "pupil-centricity" from Teacher A whose total behavior is most pupil-centered down to Teacher 0 who is least pupil-centered in her approach to teaching. Pupil-centricity is determined by the percentage of responses falling into Categories 1, 2, and 3, defined as those containing pupil-centered responses.

The product-moment correlations between categories were computed and may be found in Table II.

Summary and Conclusions

Although the sample of the study is not large, the data presented in Table II contain strong implications about relationships between various aspects or categories of teacher classroom behavior.

Some of these are as follows:

1. Teachers who were inclined to use more praise than other teachers were likely not to be as directive nor use as much reproof in their teaching.
2. Teachers whose behavior was inclined more toward the punitive seemed to use relatively little praise and were more directive than problem-structuring in contrast with other teachers.
3. It appears that teachers who used more praise than reproof were strongly inclined to be problem-structuring rather than directive in their behavior with children.
4. It should also be noted that the correlation between Category 1 + 2 (Praise) and Category 3 (Problem-Structuring) is .48 while the correlation

---

*X is a Category 6 response.

between Category (1+2) - (6+7) and Category 3 is .74. (Although tests of significance between correlation coefficients having a common variable are not entirely satisfactory, Hotellings test (2) seems to be the only practical way of determining whether the difference is significant. When this test was used a significant difference at the .01 level was found. The conditions of Hotellings test are difficult to meet and they were not met here. But as stated before there is no other test available.) These results suggest that for the teacher who had a problem-structuring teaching approach, whatever the amount of praise she may use, there will be a much higher portion of her total verbal classroom responses falling into the praise category compared to those responses she may use to chastise her students.

5. There seems to be a high negative relationship between problem-structuring and directive teacher behavior. It would seem that when a teacher made much use of one approach she made comparatively little use of the other.

The results of the study suggest that a replication of the study with a larger sample might well be worth while. It would seem desirable that the sample include teachers of all grade levels in the elementary school and perhaps also those at the secondary level.

Certainly, the findings of this study within the limits of the sample tend to confirm assumptions that have commonly been made about the classroom behavior of teachers.

REFERENCES

1. Brown, George I. A Study of the Relationship Between Classroom Climate and Learning, as Demonstrated by Competency in Reading and Arithmetic of Third Grade Pupils. Unpublished Doctoral Dissertation. Harvard University, 1958.
2. Tate, Merle W. Statistics in Education. New York: MacMillan, 1955, 467-468.
3. Withall, John "The Development of a Technique for the Measurement of Social-emotional Climate in Classrooms," Journal of Experimental Education, XIX (March 1949), 347-361.
4. Withall, John "The Development of the Climate Index." part of "Experimental Research Toward a Theory of Instruction," Thelen, H. A. (Editor). Journal of Educational Research, XLV (October, 1951). 96-98.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

THE PREDICTION OF TEACHING PERFORMANCE

DAVID L. COLE
Occidental College

THE PURPOSE of this paper is to report the findings of a five-year study conducted to determine if material obtained from certain measures of personality was useful in predicting success in elementary and secondary teaching. The criteria of success were two-fold: the ratings given a teacher by a classroom observer, and the rating given a teacher by his principal. Findings relative to the first criterion will be discussed here.

The test data used were the results from the group form of the MMPI and the group Rorschach, both tests taken during the credential candidate's undergraduate training. The predictive efficacy of these data was compared with that of data otherwise used in the selection of applicants for the credential program.

Preliminary Steps and Method

Two years prior to the full-scale follow-up, 60 of the teachers on whom test data were available were observed in the classroom by Dr. John Gowan. He rated these teachers, using a scale developed by Ryans (3, 4). The ratings given by Gowan were then related to the previously obtained test data. Five characteristics were identified which seemed related to high or low Gowan ratings. An elevation of the Schizophrenic scale of the MMPI seemed associated with low ratings. The presence of quantitatively and qualitatively good M scores on the Rorschach was associated with high ratings. The better rated teachers were more apt to have a dominance of FC over CF scores on the Rorschach. An F percent exceeding 70 percent was associated with low ratings. Subjective expressions of anxiety, exemplified by referring to the stimuli as "frightening," "threatening," "about to pounce," etc., appeared more commonly in the protocols of the lower rated teachers.

Translating the above into tentative descriptions, it appeared that when the personality test materials suggested more than ordinary difficulty in empathy, emotional control, or heightened rigidity or anxiety, the teacher in question was a poorer than average risk, as judged by a classroom observer.

Using these empirically derived cues as a basis, predictions were then made, as to the ratings which would be received by a sample of over 200 teachers. These predictions were made without reference to the identity of the teacher, and the aim of the prediction was to place each teacher in a quintile, as compared to teachers of equivalent experience, and teaching at a similar grade level. One hundred and forty of these subjects proved to be accessible to observation, and the data derived from such observations constitutes the basis for the balance of this paper.

At the time of the observations, the observer did not know what predictions had been made for the teacher in question. None of the observees covered in this phase of the report had been seen earlier by Gowan. Both Gowan and the present observer rated another sample of teachers simultaneously, however, to maximize the likelihood that the new observer would interpret the Ryans scale as Gowan had done in the pilot observations.

In the major follow-up, each teacher was visited twice, except in exceptional circumstances. Each visit was for about forty minutes. If, after two visits, the observer was unsatisfied as to the consistency of the teacher's performance, a further visit was arranged. In each case the teacher was rated by the Ryans scale, and the final score for a given teacher was the mean score of the two totals. No teacher was visited who did not consent to the visit. The teachers were interviewed, as were their principals. A single observer made all contacts.

Results: Preliminary Considerations

It was necessary to divide the 140 subjects into a number of sub-groups, in order that a teacher might be compared with a peer sample. First-year teachers generally received lower ratings than experienced teachers, and so were treated as a separate sub-group. Teachers working at the inter-

---

* The author wishes to acknowledge the Rosenberg Foundation, San Francisco, California, which provided financial support for this study.

mediate grade levels were generally rated above teachers at both younger and older levels. This was perhaps an artifax of the Ryans scale, as interpreted in this study. The scale appears to be built around two focal areas: the area of the handling of subject matter, and the area of interpersonal relationships between teacher and pupil. Teachers working with very young pupils tended to score high in this latter area, but found it difficult to score well in the subject matter area, while the converse was often true at the upper grade levels. Teachers of the intermediate grades seemed best situated to score in both areas. Hellman (1) has suggested that the intermediate grades are apt to produce a teacher's optimum performance because children in those years are least threatening to the teacher's defenses.

In any event, teachers were compared with their own approximate grade level peers, rather than with the sample as a whole. In each case, the quintile placement received by such comparison was contrasted with the predicted quintile. The quintile scores obtained from observation were adjusted to match the proportions of quintiles predicted for the sub-sample within which the teacher was being compared. Thus for the total sample, 20 percent had been predicted for the top quintile, but in a given sub-group, it might turn out that 30 percent of those included had been predicted for the top quintile. In such an event, the top thirty percent of the sample, as rated by observation, were considered to constitute the top quintile. This avoided inaccuracy of prediction which would have otherwise arisen merely from the necessity of forming the sub-groups after the initial blind predictions had been made.

Results: Relation of Observer Ratings to Conventional Selection Criteria

The teachers seen in this study had been admitted to credential candidacy on the basis of several criteria. These included their lower division grade record, their entrance examination scores, the evaluations made of them by faculty members, and an evaluation made by the Dean of Men or Dean of Women, as appropriate. After the observer ratings had been obtained, the personnel files of the subjects were studied, and their status with reference to these selection criteria determined. Where possible, these were translated into quintile rankings, and these compared with the quintile placement from observer ratings. A composite quintile was also determined, based on the several selection measures, and this also compared with the quintile placement based on classroom observation. These results are summarized in Table I.

It will be noted that a non-significant correlation was found between observer ratings and the lower division grade record, the various measures from the ACE, and the composite predicted quintile from the selection measures. The ratings by the personnel deans could not be handled in this manner and are discussed later.

TABLE I

CORRELATION OF OBSERVER RATINGS WITH PREDICTIONS FROM VARIOUS CRITERIA

Criterion	r	P.E.
Lower division grades	.19	.06
A.C.E. Q	.19	.06
A.C.E. L	-.09	.06
A.C.E. T	.01	.07
Faculty Ratings	.26	.08
Composite from above	.15	.06
Personality test data	.65	.03

A special comment must be made with reference to the usefulness of the rating forms completed by faculty members familiar with the applicant. The form used was a very simple one, listing seven desirable "traits," with the request being that the faculty member identify each as present, or absent. The rater was also given the option of pleading ignorance. Inspection showed that negative ratings were extremely rare, although ignorance was often admitted. Fully one third of the subjects had been rated positively on each of the seven characteristics by each of the three required faculty raters.

To evaluate the usefulness of these ratings, the sample was divided into two parts. One part consisted of that third for whom all ratings had been positive. A distribution was made of the quintile placements, based on classroom observation of these subjects, and this distribution compared, by a Chi Square test, to the total distribution. No significant difference was found, Chi Square being .19. Thus it appears that when all faculty raters were uniformly positive in rating an applicant, this had no predictive significance at all.

From this it was hypothesized that uniformly positive ratings meant either that the applicant had been very careful in the selection of his faculty references, or that the persons completing the forms had been careless or non-discriminating. It was further hypothesized that this would be less true of those applicants for whom the ratings were not unanimously positive. A distribution was made, therefore, of the ratings of this group, quintile placements assigned, and these compared with the observation results. The resulting r of .26, reported in Table I, approaches significance, and may suggest that faculty references, if made with more care and discrimination, may be a useful predictor of future success in teaching. It may also suggest that when a faculty rater pleads ignorance, he is, in fact, hesitating to express a negative feeling.

The ratings by the personnel deans could not be satisfactorily translated into quintile placements, since they consisted of open-end comments. It was possible, however, to translate these into the three categories of "favorable," "neutral," and "unfavorable or hesitant." When this was done, it was then possible to consider the frequency with which a person identified as more than ordinarily promising by the personnel dean was in fact rated by the observer as below average, as well as the frequency of the converse of this situation. This was termed a "reversal," and was a process which could also be applied to the other criteria. Thus any candidate predicted for one of the top two quintiles who was rated in one of the bottom two constituted a "reversal," as would the converse of this situation. The percentage of such reversals for each standard criterion measure is described in Table II. It will be noted that by this measure, the evaluations by the personnel deans are not more valuable than the other conventional criteria.

TABLE II

PERCENTAGE OF "REVERSALS" BETWEEN RATINGS AND PREDICTIONS BASED ON VARIOUS CRITERIA

Criterion	Percentage of Reversals
Lower division grades	32
A.C.E. Q	31
A.C.E. L	41
A.C.E. T	38
Faculty Ratings	25
Ratings by Personnel Deans	33
Composite from above	33
Personality test data	11

Results: Relation of Observer Ratings to Predictions from Personality Test Materials

These relationships have been noted in Tables I and II, at the bottom of each table. It will be noted that a significant correlation was found, and that the percentage of reversals was far below that found by other criteria.

Discussion

The fact that the established criteria for selection of credential candidates proved ineffective in predicting their eventual success is not sufficient to warrant discarding these criteria. The present data do not of course speak to the validity with which these established criteria identified those eliminated from the credential program. They do suggest, however, that once a candidate met the established criteria, these criteria were of no further usefulness in predicting his success in teaching. By contrast, the personality test materials do appear useful. Inspection of the materials, after the results had been determined, suggested that the two most useful cues within this material were the quantity and quality of the M scores on the Rorschach, and the FC:CF balance on the same instrument. Teachers who had fewer than three main M scores, or who gave M scores which were consistently associated with derogatory specifications appeared to have difficulty in anticipating and understanding the feelings and needs of their students, and difficulty in gaining their cooperation. Teachers who showed a preponderance of CF scores over FC scores showed evidences of impulsivity, and difficulty in expressing controlled, forceful emotions which maintained classroom discipline, without resorting to open hostility or aggression. They seemed handicapped by their own impulsivity when called upon to meet the impulsivity of the children.

This paper has not dealt with the ratings given by the principals. This is due to the fact that these ratings were so skewed to the high end of the scale as to make comparison with the predicted ratings very difficult. In a further study, now under way, an attempt is being made to correct this situation.

Summary

In a study attempting to determine if material in the group Rorschach and MMPI could be used to predict success in teaching, as determined by the ratings of a classroom observer, significant relationships were found. Prediction based on the conventional selection criteria was not significantly above chance.

REFERENCES

1. Hellman, Else. "Psychoanalysis and the Teacher," in J. D. Sutherland (Ed.), Psychoanalysis and Contemporary Thought. New York: Grove Press, 1959, pp. 58-77.
2. Klopfer, Bruno; Ainsworth, Mary; Klopfer, Walter, and Holt, Robert. Developments in the Rorschach Technique, Vol. I, New York: World Book Company, 1954.

3. Ryans, David G. "A Study of Criterion Data," Educational and Psychological Measurement, 1952, 12, 333-44.

4. Ryans, David G. "The Investigation of Teacher Characteristics," Educational Record, 1953, 34, 371-96.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

ANALYSIS OF ATTITUDE INVENTORY DATA ON MINNESOTA VOCATIONAL AGRICULTURE TEACHERS

GORDON I. SWANSON, University of Minnesota and
LOYAL W. JOOS, Baltimore County Schools

Purpose of the Study

THE PRIMARY purpose of this study is to secure basic information concerning the attitudes among Minnesota Vocational Agriculture instructors, as measured by the Sundet Attitude Inventory.[1]

A secondary purpose of the study was to determine the correlations (if any) between attitude scores and each of five categories of additional information:

1. Total years of teaching experience.
2. Years of experience in present school.
3. Graduate work expressed in quarter credits.
4. Percent of teaching experience years that teacher had a Young Farmer instructional program.
5. Percent of teaching experience years that teacher had an Adult Farmer instructional program.

It might be expected that there would be some relationship between years of experience and attitude score; and it is also reasonable to assume a strong relationship between a high attitude score and a high incidence of Young Farmer instruction expressed as a percentage of total instructional time since the Young Farmer program in most cases is optional for both teacher and student.

Procedure

The Sundet Attitude Inventory, together with an answer sheet and an information sheet, with a covering, explanatory letter, was sent to each of the 319 Vocational Agriculture Teachers in Minnesota in 1958. After one follow-up request, 304 completed, usable returns were available. This number constituted 95.3% of the population of agriculture teachers in the state of Minnesota.

The returns were tabulated, measures of variation and central tendency were computed, and a table to translate raw to percentile scores was prepared. Relationships between attitude scores and other measures were determined.

The following statistics were calculated:

a. Attitude Scores (from grouped data)
  Mean score = 108.32
  Standard deviation = 22.3
  Median score = 106.75
  Range in scores = 111
    Low score = 55
    High score = 166
  Interquartile range = 32.38
    Mode = 113

b. Total Years of Experience
  Mean = 8.8
  Range = 38 years
  C 50 = 6.8 years

c. The average teacher had been in the present school system 5.45 years and he had accumulated 19.7 quarter credits of graduate work. No averages were calculated on the Young Farmer and Adult Farmer Programs.

On the basis of group data on attitude scores, Table I was prepared. The purpose of this table is to aid in interpreting such scores in the future.

The overall Pearson Product moment correlation for attitude scores and experience value was .02. This value appeared to be smaller than expected. Since the scattergram indicated some linear relationship over the first ten years of experience, and less over the entire range of experience, a second r was calculated, using only scores of teachers with ten years or less of experience. The resultant value, $r = .182$, is significant at the .01 level for $N = 230$. A regression line fitted to the plot of attitude scores of these teachers, by method of least squares, has the formula

$$Y = 108 + 1.4\,(x-5) = 1.4\,x + 101$$

The correlation coefficient between attitude score and Years of Experience in Present School was calculated to be .069. This value is too low to be significant. Moreover, there appears to be no rationale which might explain any correlation if it were significant.

---

[1] Sundet, Stanley, "An Attitude Inventory for Teachers of Vocational Agriculture." Unpublished Ph. D. Thesis, Graduate School, University of Minnesota. 1955. 97 pp. Library, University of Minnesota.

TABLE I

CONVERSION OF ATTITUDE RAW SCORES TO PERCENTILES

Raw Score	Percentile
156-160	100
151-155	98
146-150	97
141-145	95
136-140	92
131-135	86
116-130	81
111-115	66
106-110	55
101-105	48
96-100	39
91-95	31
86-90	21
81-85	15
76-80	12
71-75	7
66-70	3
61-65	2
56-60	1
51-55	0

TABLE II

YOUNG FARMER PROGRAM

Attitude Score	Have had a Young Farmer Program	Never had a Young Farmer Program	Total
100+	139	48	187
99-	65	50	115
Total	204	98	302

Chi-square value = 9.996, significant at 1% level.

TABLE III

RELATIONSHIP BETWEEN ATTITUDE SCORE AND TOTAL VOCATIONAL AGRICULTURE PROGRAM

Attitude Score	No Young Farmer or Adult Program	Adult Program Only	Both Young Farmer and Adult	Total
108+	5	21	116	142
107-	20	50	92	162
Totals	25	71	208	304
Chi-square	7.1495	8.3667	6.8555	22.3717
Level of Significance	5%	5%	5%	1%

A scattergram plot of graduate credits versus attitude score seemed to show a trend toward higher scores for those with more graduate work. Closer examination of the plot suggested, however, that the increase might well be due to the fact that, in general, those teachers with more graduate work would be those with more years of experience. To partially exclude the effect of experience, a contingency table was set up on the 71 teachers that had from one to three years of teaching experience.

The resultant Chi-square value was 5.69. This would be equalled or exceeded 15% of the time by chance. The hypothesis that there is no significant difference between attitude scores of teachers categorized according to graduate credits earned would not be rejected on the basis of these data for this group.

In order to determine the relationship between attitude score and the presence or absence of a Young Farmer program, a contingency table and Chi-square test of independence was prepared, as shown in Table II.

A reasonably direct relationship between attitude score and the presence or absence of a Young Farmer program might be assumed since, in most cases, a teacher will have such a program as an optional, rather than a required part of his total program. The actual distribution of teachers in the above table could have occurred by chance less than one out of a hundred times, if there were no relationship between the Young Farmer program and attitude. Accordingly, the hypothesis of no relationship was rejected.

A calculation of Pearson Product moment correlation between Attitude score and the extent of experience with Young Farmer programs yields a value of $r = .26$.

A more complete tabulation of the relationship between Attitude score and the total program carried on by agriculture teachers is shown in Table III.

We may safely reject the hypothesis that there is no relationship between Attitude score and type of program carried out as it pertains to Young Farmer and Adult programs.

Summary

This study presents certain basic information concerning attitude toward teaching by Minnesota Agriculture teachers. It also provides a conversion scale for interpreting percentile scores from raw scores based on attitude measurements of 304 teachers.

A small but statistically significant relationship is established between average Attitude scores and years of teaching experience with an appropriately fitted regression line for experience up to ten years.

In general, the evidence of the study indicates that a significant relationship exists between attitude score and the type or program conducted as it relates to Young Farmer and Adult Farmer instruction. In most schools, the Young Farmer and Adult Farm programs are optional to the teacher and they are invariably optional to students. It may be said, therefore, that on the average, the relationship of Attitude scores to the total program provides an effective system of classification as it pertains to the teaching of these enrollments in Minnesota.

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

A FOLLOW-UP STUDY OF TEACHING CANDIDATES

J. C. GOWAN, CARITA CONNER, AND PHYLLIS KENNEDY*
San Fernando Valley State College
Northridge, California

THIS REPORT gives details of a follow-up study of 329 teaching candidates who completed student teaching and whose school of employment was known two years later. Although this group was a selected sample, there is no reason to believe that it is in any way unrepresentative of the teaching candidates generally enrolled in the Education Division of California State Colleges. All of the 329 subjects were enrolled at the Vermont campus of the Los Angeles State College between 1953 and 1956. A breakdown by credential and sex reveals the following figures:

Sex	Elementary	Secondary	Special	Total
Men	41	116	26	183
Women	83	41	22	146
Total	124	157	48	329

Problem

The primary purpose of this research was to study the relationship between training ratings and tests of student teachers with the ratings of principals of these teachers several years later.

Method

Measures used were of three types: 1) training ratings, 2) tests, 3) principals' ratings. Training ratings were obtained from cumulative records and the training consensus was obtained by pooling a) the ratings of the master teacher in the candidate's student teaching, b) similar ratings of the supervising college instructors, and c) a placement office rating. Test and other results on many of the candidates included 1) raw score on the Minnesota Teacher Attitude Inventory, 2) percentile score on an experimental teacher prognosis test, 3) grades in the introductory course in Educational Psychology handled by the senior author, and 4) a class evaluation of the candidate in the same course.

Since the field ratings were within the control of the study, a specially prepared questionnaire was designed to elicit as much information as possible from the principals and to assist in interpreting the internal consistency of their report. The final instrument, (available upon request) consisted of a one page sheet involving three types of information: a) an evaluation of ten categories of the overall performance of the teacher, b) a fifteen item forced-choice selection of one of four descriptive adjectives. Of these, according to a previous study[1], one was valent and positive, one ostensibly positive but actually non-discriminating, one valent negatively, and the escape phrase "none apply", and c) a list of some twenty-one statements about the teacher, any or none of which might be checked. The instrument was sent to principals of 285 subjects, and usable replies were received from 222. Training consensus statistics on the 107 cases from which no usable reply was received indicated no significant difference in the mean or standard deviation from the usable sample. A field rating consensus was constructed by combining the over-all categorization of the teacher's performance by the principal (which was heavily weighted), with a score summing the positive adjectives and phrases checked minus the negatives. Tests of reliability between the various training ratings, (see Table I) yielded correlations over .7 so that consistency was satisfied, and it may be presumed that the same variables are being assessed by different methods.

Results

The principal results of the study, namely the correlations between training and field ratings are given in Table I. There is significant correlation of .28 between training and field consensus ratings. The largest correlation between a training statistic and field consensus is between class evaluation and field consensus where the correlation reached .37. While class grades had some relation to field consensus, neither test used showed any relationship. In general it may be said that the way in which teacher candidates are seen by their fellows is the best

---

* All footnotes will be found at end of article

TABLE I

CORRELATIONS OF MEASURES IN TEACHER FOLLOW-UP STUDY

	TRAINING CONSENSUS				FIELD RATINGS			
Measure	N	r	S.E.	Signif.	N	r	S.E.	Signif.
TRAINING CONSENSUS	100	.72	.05*	.01	226	.28	.06	.01
FIELD RATINGS	226	.28	.06	.01	100	.76	.05*	.01
Minnesota Teacher Attitude Inventory	86	-.09	.12	---	52	-.01	.14	---
Experimental Personality Test	156	.14	.08	---	119	.14	.10	---
Grades in Education Class	121	.15	.09	---	91	.23	.10	.05
Class Evaluation in Same	70	.48	.09	.01	44	.37	.14	.01

* Reliabilities of main variables are compiled as follows: for training consensus, T(M) is correlated against T(S), for field ratings, F(1) is correlated against F(2). Both give minimal estimates of reliability.

predictor of the way they are going to be seen both by training authorities and field raters later.

Discussion

It should be noted that in a study of this kind we are not comparing excellent teachers with candidates who are poor enough to be screened, but instead with those who are successful in student teaching but who have some initial difficulty in the school system. This is a much narrower range, and makes high demands on the accuracy of our measuring instruments. Some of these initially poor teachers may improve significantly with time. It is perhaps too much to expect that we can predict accurately at the training level in areas of competency in which we often do not train.

When control groups, such as provisional teachers, have been used in follow-up studies, it has sometimes been found that factors which discriminate between criterion and control groups appear to be outside the control of the training institution. To a large extent the worth of the individual teacher is decided before, not after, she enters training[2]. As another study puts it: "An effective teacher is the social legacy of a good home"[3]. We do need continuing research to find out how training institutions can simulate such a climate which will induce desirable attitudinal changes toward the profession and as a result toward children.

FOOTNOTES

* J. C. Gowan is Associate Professor of Education, Carita Conner is Placement Supervisor and Phyllis Kennedy is Staff Psychometrist at San Fernando Valley State College.

1. Gowan, J. C. "The Use of the Adjective Check List in Screening Teaching Candidates," Journal of Educational Research, 49:9:663-672, (May, 1956).
2. Gowan, J. C. "A Comparison Between Women Teachers With Regular Versus Emergency Credentials," Educational Horizons, 32:134-138, (Winter, 1953).
3. Gowan, J. C. "A Summary of the Intensive Study of Twenty Highly Selected Elementary Women Teachers," Journal of Experimental Education, 26:2:115-124, (December, 1957).

JOURNAL OF EDUCATIONAL RESEARCH
(Volume 54, Number 9, May 1961)

EVALUATION OF THREE METHODS OF TEACHING READING, SEVENTH GRADE

FREDERIC R. WALKER
West Sayville, New York

RECENT INTEREST in teaching reading beyond the elementary level has resulted in the introduction of a variety of new methods and materials. It was the purpose of this study to evaluate and compare the Science Research Associates "Reading Laboratory" and a more strictly individualized program with a rather conventional approach. Student progress was measured by comparing comprehension and vocabulary achievement test scores before and after instruction, in each of the three classes. In addition, some measurement of the progress of the lower half of each class was determined. This seemed of particular interest as there was probably a larger gap between their achievement and capacity.

The eighty-six children in this study were members of three, heterogeneously grouped, classes in Sayville, New York. The children were all seventh graders, largely from middle class homes. There were no significant differences among the three classes in size, intelligence, reading grade level, or chronological age. According to the school records, both the average reading grade level and the intelligence quotient fell between the 55th and 60th percentile. The entire study was conducted over a period of six weeks for one period (forty-five minutes) of instruction each day.

In order that the instruction would be relatively free of holidays, the program was put into operation during January and February, 1959. The procedure used in each of the three classes was as follows:

7-1, (Experimental Group), was a class in which a variety of materials were used in an attempt to meet the individual child's particular difficulty. Children without any particular difficulty were given training in critical reading, especially analyzing advertisements. About one half of the time was spent in doing these activities, which were either teacher made, or carefully selected parts of different workbooks. The aim was to have the child develop only those particular skills in which he was weak, and not to waste time doing repetitious drill in skills that he already knew. The remainder of the time was spent in doing vocabulary study and reading comprehension (and speed) activities.

7-2, (SRA Reading Group), received instruction only as directed by the SRA "Reading Laboratory". The only modification was made during the last two weeks of the study. The activities were accelerated; instead of using the "Reading Laboratory" every other day, it was used every day. This was done so that continuous reading instruction could be maintained.

Both 7-1 and 7-2 used a form of individualized instruction. However, the individualized instruction in the SRA "Reading Laboratory" is merely placing the child in activities at his grade level, whether he needs them or not. There is no consideration for individual difficulties, although the most recent edition of the "Reading Laboratory" teacher's manual has taken steps to correct this.

7-3, (Control Group), followed a pattern which resembled a more conventional approach to teaching reading. One seventh grade textbook, and the accompanying workbook, were used with the entire class. The teacher spent time developing the child's interest in the stories, and in leading discussions after the silent reading. No provision for individualized instruction was provided, nor was any supplementary reading material employed.

Two methods of evaluating the three classes were used:

1. The progress of comprehension and vocabulary in each class, and the lower half of each class, was measured by the Diagnostic Reading Tests, Survey Section, Lower Level, Frances O. Triggs, (New York, Committee on Diagnostic Reading Tests, Inc. 1952). Forms A and B were used in determining the pre and post instruction scores.
2. The amount of time required by the teacher outside of class in preparing and planning was recorded.

Although the reading grade level for each of the three classes averaged between the 55th and 60th

TABLE I

COMPREHENSION AND VOCABULARY PROGRESS BEFORE AND AFTER INSTRUCTION

Group	Percent Increase		Fisher's t test	
	Comprehension	Vocabulary	Comprehension	Vocabulary
Entire Class:				
7-1 (Exp)	8.3	13.4	1.48	2.05
7-2 (SRA)	11.1	9.4	1.68	1.40
7-3 (Con)	6.2	7.1	.96	1.63
Lower half of each class:				
7-1 (Exp)	24.2	20.4	2.43*	2.45*
7-2 (SRA)	18.9	17.1	2.12*	2.09*
7-3 (Con)	8.3	5.6**	.97	.94

*Significant at the .05 level of confidence

**Due to the higher starting level, this score was prorated. Actual increase was 4.5%.

percentile on the school's standardized tests, the starting level on the *Diagnostic Reading Tests* was somewhat lower. The average starting levels of the three classes fell between the 35th and 39th percentile in comprehension, and between the 50th and 56th percentile in vocabulary. The lower halves of the classes all started at the 20th percentile in comprehension. In vocabulary the lower halves of 7-1 and 7-2 started at the 23rd percentile; 7-3 started about six points higher. This would tend to distort the per cent increase of 7-3, however this distortion was not significant.

Conclusions

Table I shows the per cent of increase, and the significance of the increase, in comprehension and vocabulary, as calculated from the pre and post test raw scores.

Although there was a difference in the per cent increases of the entire classes, the differences between any two classes in the final test scores were not significant. The progress within the three classes varied, but none of the classes improved significantly. However, it should be noted that 7-1's vocabulary progress fell just short of being significant at the .05 level of confidence, (Fisher's t test was used in determining significant progress).

In measuring the progress of the lower halves of each class, substantial and significant progress was made in both 7-1 (Experimental Group) and 7-2 (SRA Reading Group). Vocabulary and comprehension abilities made much greater improvement among the poorer readers.

Additional data revealed that some progress was made by a few pupils in 7-3 (Control Group) who started instruction at or just above grade level (in reading). Other data revealed that the progress in 7-2 was quite even; each pupil made a small degree of improvement. In 7-1 some children made very large increases, others made almost no progress. In comparing the time spent out of class by the teacher, there is little question that the 7-2 class required the least preparation. 7-1 required the most time, but much of the time was spent in preparing materials that may be used again and again. The methods used in a program similar to 7-1 are subject to constant improvement and revision, whereas the methods of 7-2, the SRA "Reading Laboratory", are determined by the publisher.

A likely reason for the greater improvement in 7-1 and 7-2, than in 7-3, was due to the individualized instruction. It is possible that the better readers in 7-1 could have made greater improvement if they had received specific comprehension training, instead of critical reading which was not measured by the tests. The poorer progress displayed by 7-3 was probably due to the one level of materials, which met the difficulties of only a small part of the class.

In a final analysis, the results of this study tend to support an individualized program. Since significant progress was made with the SRA "Reading Laboratory", it should be given a definite part in the reading program. This would be very true of a class where the teacher is either inexperienced or over burdened with other duties. An inexperienced teacher would have much difficulty preparing individual activities, and an over-burdened teacher would have little time to develop individual exercises. However, the results of this experiment favored the program of 7-1 slightly. 7-1 was especially good because it provided for future development and a more precise meeting of the child's difficulty. Also, 7-1 has a more active part for the teacher which brings in a human element, that even the finest materials cannot replace.

BUREAU OF EDNL. & PSYL. CAL
LIBRARY